



Office de la propriété  
intellectuelle  
du Canada

Un organisme  
d'Industrie Canada

Canadian  
Intellectual Property  
Office

An Agency of  
Industry Canada



*Bureau canadien  
des brevets*  
Certification

La présente atteste que les documents  
ci-joints, dont la liste figure ci-dessous,  
sont des copies authentiques des docu-  
ments déposés au Bureau des brevets.

*Canadian Patent  
Office*  
Certification

This is to certify that the documents  
attached hereto and identified below are  
true copies of the documents on file in  
the Patent Office.

#  
20  
hnda  
4-24-03

Specification and drawings, as originally filed, with Application for Patent Serial No:  
2,332,733, on February 20, 2001, by **DENNIS G. CVITKOVITCH, PETER C.Y. LAU  
AND YUNG HUA LI**, for "Signal Peptides, Nucleic Acid Molecules and Methods for  
Treatment of Caries."

*Bray Paulhus*  
Agent certificateur/Certifying Officer

January 27, 2003

Date

Canada

(CIPQ 68)  
04-09-02

OPIC  CIPO

## Abstract

The invention relates to a compound that competitively inhibits binding of CSP to *S. mutans* histidine kinase. The compound is preferably a peptide or an antibody. The compound is preferably a derivative of [SEQ ID NO:2], a fragment of [SEQ ID NO:2] or a derivative of a fragment of [SEQ ID NO:2].

## **SIGNAL PEPTIDES, NUCLEIC ACID MOLECULES AND METHODS FOR TREATMENT OF CARIES**

### **5 FIELD OF THE INVENTION**

The invention relates to compounds and methods that inhibit or disrupt microbial biofilms involved in infections in man and animals and in biofouling of surfaces susceptible to microbial accumulation.

### **10 BACKGROUND OF INVENTION**

Bacteria often attach and accumulate on surfaces, enabling them to resist removal and killing by mechanical and chemical means. This can result in persistent and chronic infections and fouling of devices that are in contact with liquids containing the colonizing bacteria.

15 Controlling bacterial biofilms is desirable for almost every human enterprise in which solid surfaces are introduced into non-sterile aqueous environments. US Patent No. 6,024,958 describes peptides that attempt to control biofilm formation by preventing bacterial adherence to teeth. In addition to occurrence in dental caries, medical  
20 examples of biofilm growth include cases involving indwelling medical devices, joint implants, prostatitis, endocarditis, and respiratory infections. In fact, the Centers for Disease Control and Prevention (CDC; Atlanta, GA) estimate that 65% of human bacterial infections involve biofilms. Non-medical examples of biofilm colonization are water and beverage lines, cooling towers, radiators, aquaculture contamination, submerged pumps and impellers, hulls of commercial, fishing and military vessels and  
25 literally every situation where biofouling occurs. The potential benefits of basic research focused at biofilm physiology and genetics with the ultimate goal of controlling surface-mediated microbial growth are limitless.

Interest in the study of biofilm-grown cells has increased partly because biofilm growth provides a microenvironment for cells to exist in a physical and physiological  
30 state that can increase their resistance to antimicrobial compounds and mechanical forces (reviewed in Costerton and Lewandowski, Adv Dent Res, 11:192-195). Growth in biofilms can also facilitate the transfer of genetic information between different species (Christensen *et al.* Appl Environ Microbiol, 64:2247-2255). Recent evidence suggests that biofilm-grown cells may display a dramatically different phenotyp when compared

with their siblings grown in liquid culture. In some, this altered physiological state has been shown to result from gene activation initiated by contact with surfaces (Finlay and Falkow. Microbiol Molec Rev, 61:136-169) or from signal molecules produced by the bacteria allowing them to sense the cell density (quorum sensing) (Davies *et al.* Appl Environ Microbiol, 61:860-867). Biofilms may also act as 'genotypic reservoirs', allowing persistence, transfer and selection of genetic elements conferring resistance to antimicrobial compounds.

*Streptococcus mutans* is the principal etiological agent of dental caries in humans. None of the known types of *S. mutans* antibiotics has satisfactorily controlled caries. There is a need to identify new ways to control *S. mutans* induced caries.

## SUMMARY OF THE INVENTION

The invention relates to a compound that competitively inhibits binding of CSP to *S. mutans* histidine kinase. The compound is preferably a peptide or an antibody. The compound is preferably a derivative of [SEQ ID NO:2], a fragment of [SEQ ID NO:2] or a derivative of a fragment of [SEQ ID NO:2].

## BRIEF DESCRIPTION OF THE DRAWINGS

Preferred embodiments of the invention will be described in relation to the drawings in which:

Figure 1

Schematic layout of the arrangement of the genetic locus encoding the signal peptide precursor (ComC), the histidine kinase (ComD) and the response regulator (ComE). Note that this arrangement is different from other loci in related streptococci for the following reasons: a) The comC gene is transcribed from its own unique promoter, unlike the genes thus far described in other streptococci that are arranged in an operon-like cluster with the comC/DE genes being transcribed from a single promoter. b) The comC gene is separated by 148 nucleotides from the comD gene.

Figure 2

Shows the nucleic acid molecule that is [SEQ ID NO:1]. In a preferred embodiment, the figure shows CSP (competence signal peptide). Nucleotide sequence of the locus. Figure 2 also shows histidine kinase sequences and response regulator sequences.

5    Figure 3

Sequence of the deduced amino acid sequence of the signal peptide, histidine kinase, and response regulator.

Figure 4

- 10   The deduced amino acid sequence of the signal peptide precursor and its predicted cleavage site, following the glycine-glycine signal.

Figure 5

- 15   Shows the peptide that is [SEQ ID NO:2]. The synthetic signal peptide that is effective at inducing competence, biofilm formation and acid tolerance in *Streptococcus mutans*.

Figure 6

- 20   The natural activity of the signal/receptor system functioning in vitro in model biofilms as determined by the ability of various strains of *S. mutans* to accept donor plasmid DNA conferring erythromycin resistance.

Figure 7

- 25   Table illustrating the effect of synthetic peptide on genetic competence in *S. mutans* cells.

Figure 8

List of the primers used to amplify the genes or internal regions of the target genes by polymerase chain reaction (PCR) for subsequent sequencing or inactivation.

30   Figure 9

ComCDE local region.

Figure 10

The comX DNA sequence, protein sequence, and the comX gene local region with 100bp included both upstream and downstream (promoter is upstream).

Figure 11

- 5 The comA and comB nucleotide and amino acid sequences. ComA and ComB are the components of the CSP exporter.

Figure 12

- 10 Illustrates the effect of synthetic peptide on acid resistance tolerance in *S. mutans comC* deficient cells.

#### DETAILED DESCRIPTION OF THE INVENTION

- 15 In some Gram-positive bacteria (including *Streptococcus mutans*), when a specific histidine kinase receptor located in the cell membrane is disrupted, the cells become ineffective at developing a biofilm. The cells growing in this biofilm environment use a small peptide signal molecule to activate the receptor in surrounding cells, thereby communicating the message to form a biofilm. This same signal peptide and histidine kinase are also involved in the induction of genetic competence, the cell's ability to take up and incorporate DNA from its extracellular environment, as well as that of acid tolerance, the cell's ability to survive pH levels as low as pH 3.0. A mechanism that blocks the signal molecule from activating the histidine kinase receptor molecule provides a novel method for controlling microbial biofilms, either alone or in combination with chemical or physical means.

- 25 We have identified a genetic locus in *S. mutans* consisting of three genes that encode: 1) a peptide precursor that is processed during export into a secreted 21-amino acid peptide (CSP); 2) a histidine kinase that acts as a cell surface receptor activated by the peptide; 3) a response regulator that activates a number of other genes involved in genetic competence, biofilm formation, and acid tolerance of *S. mutans*. These properties have been attributed to the bacterium's ability to cause dental caries. Inactivation of any of these three genes or impairment of interaction or activity of any of their encoded proteins will disrupt the bacterium's ability to take up foreign DNA, form biofilms, and tolerate acidic pH.
- 30

*Streptococcus mutans* is a resident of the biofilm environment of dental plaque, a matrix of bacteria and extracellular material that adheres to the tooth surface. Under appropriate environmental conditions populations of *S. mutans* and the pH of the surrounding plaque will drop. *S. mutans*, being among the most acid tolerant organisms residing in dental plaque, will increase its numbers in this acidic environment and eventually become a dominant member of the plaque community. This situation eventually leads to dissolution of the tooth enamel, resulting in the development of dental caries. We control the accumulation and acid tolerance of this bacterium to make it less able to cause caries. We accomplish this by using inhibitors of an extracellular signal peptide that promotes the expression of genes involved in *S. mutans* biofilm formation and acid tolerance. The invention includes compounds that inhibit the action of the peptide. These inhibitors can include peptides, antibodies, or other agents that specifically inhibit the activation of the histidine kinase and the family of genes activated as a result of the histidine kinase activation by the signal molecule. Inhibitors include: modified structures of the peptide where amino acids are removed from the N- and/or COOH terminal of the peptide and/or substitutions of internal amino acid residues. We delete, one, two to 5, 6 to 10 and 10 to 15 amino acids from the peptide (for example at either terminal) and measure competitive inhibition of signal peptide binding to histidine kinase (1, 2, 3, 4, 5, 6, 7, 8, 9, or 10 or more amino acids are deleted and inhibition measured). Inhibitors also include antibodies raised against the 21-amino acid CSP alone or coupled to a larger molecule to increase immunogenicity.

We also test inhibitors described in (Barrett et al. Proc. Natl. Acad. Sci USA 95:5317-5322) and measure competitive inhibition of signal peptide binding to histidine kinase.

In addition to identifying the genes encoding this signaling/sensing system, we have identified and chemically synthesized a 21-amino acid peptide that promotes biofilm formation and acid tolerance of *S. mutans*. A survey of the literature and genome databases reveals that genes similar to this signal-receptor system are present in most Gram-positive bacteria, and therefore an inhibitor, or family of related inhibitors may be effective at inhibiting biofilm formation among a large group of bacteria.

The invention treats or prevents dental caries by addition of compounds that inhibit the stimulatory action of the 21-amino acid peptide on biofilm formation and acid tolerance of *S. mutans*. This is accomplished by delivery of these compounds to the biofilm and/or to incorporate these inhibitors into materials to control growth on surfaces.

This includes delivery by topical application, alone or in combination with other compounds including toothpaste, mouthwash, food or food additives.

*Streptococcus mutans* is also implicated in causing infective endocarditis. Inhibitors of biofilm formation, and hence aggregation are useful in the treatment of these bacterial infections as well.

#### **Identification and characterization of Competence Signal Peptide (CSP), Histidine Kinase (HK) and Response Regulator (RR)**

##### *Competence Signal Peptide*

The invention includes an isolated CSP from *S. mutans*. The invention also includes a recombinant isolated CSP peptide produced by a cell including a nucleic acid molecule encoding CSP operably linked to a promoter. The invention also includes an isolated nucleic acid molecule encoding a CSP. The peptide we work with is preferably chemically synthesized.

The invention includes CSP-encoding nucleic acid molecules and molecules having sequence identity or which hybridize to the CSP-encoding sequence and which encode a peptide having CSP activity (preferred percentages for sequence identity are described below) as well as vectors including these molecules. The invention also includes CSP or peptides having sequence identity (preferred percentages described below) or which have CSP activity. The nucleic acid molecules and peptides of the invention may be from *S. mutans* and they may be isolated from a native source, synthetic or recombinant. The invention includes CSP or peptides having sequence identity, which have CSP activity, as prepared by the processes described in this application.

##### *Histidine Kinase*

The invention includes an isolated HK from *S. mutans*. The invention also includes a recombinant isolated HK polypeptide produced by a cell including a nucleic acid molecule encoding HK operably linked to a promoter. The invention also includes an isolated nucleic acid molecule encoding a HK polypeptide.

The invention includes HK-encoding nucleic acid molecules and molecules having sequence identity or which hybridize to the HK-encoding sequence and which encode a protein having HK activity (preferred percentages for sequence identity are described below) as well as vectors including these molecules. The invention also



includes HK or polypeptides having sequence identity (preferred percentages described below) or which have HK activity. The nucleic acid molecules and polypeptides of the invention may be from *S. mutans* and they may be isolated from a native source, synthetic or recombinant. The invention includes HK or polypeptides having sequence  
 5 identity, which have HK activity, as prepared by the processes described in this application.

#### *Response Regulator*

The invention includes an isolated RR from *S. mutans*. The invention also includes a recombinant isolated RR polypeptide produced by a cell including a nucleic  
 10 acid molecule encoding RR operably linked to a promoter. The invention also includes an isolated nucleic acid molecule encoding a RR polypeptide.

The invention includes RR-encoding nucleic acid molecules and molecules having sequence identity or which hybridize to the RR-encoding sequence and which encode a polypeptide having RR activity (preferred percentages for sequence identity  
 15 are described below) as well as vectors including these molecules. The invention also includes RR or polypeptides having sequence identity (preferred percentages described below) or which have RR activity. The nucleic acid molecules and polypeptides of the invention may be from *S. mutans* and they may be isolated from a native source, synthetic or recombinant. The invention includes RR or polypeptides having sequence  
 20 identity, which have RR activity, as prepared by the processes described in this application.

The comA and comB nucleotide and amino acid sequences are also aspects of the invention. ComA and ComB are components of the CSP exporter. The discussion of variants, sequence identity etc. for CSP, HK, RR applies to both the full sequences  
 25 shown in the figures as well as bracketed portions of sequences (coding regions). The peptides and polypeptides may be natural, recombinantly produced or synthetic.

#### **Functionally equivalent nucleic acid molecules**

The invention includes nucleic acid molecules that are functional equivalents of  
 30 all or part of the CSP sequence in [SEQ ID NO:1]. (A nucleic acid molecule may also be referred to as a DNA sequence or nucleotide sequence in this application. All these terms have the same meaning as nucleic acid molecule). Functionally equivalent

nucleic acid molecules are DNA and RNA (such as genomic DNA, complementary DNA, synthetic DNA, and messenger RNA molecules) that encode peptides having the same or similar CSP activity as the CSP peptide shown in [SEQ ID NO:2]. Functionally equivalent nucleic acid molecules can encode peptides that contain a region having

5 sequence identity to a region of a CSP peptide or more preferably to the entire CSP peptide. Identity is calculated according to methods known in the art. The ClustalW program (preferably using default parameters) [Thompson, JD et al., Nucleic Acid Res. 22:4673-4680.], described below, is most preferred. For example, if a nucleic acid molecule (called "Sequence A") has 90% identity to a portion of the nucleic acid

10 molecule in [SEQ ID NO:1], then Sequence A will preferably be identical to the referenced portion of the nucleic acid molecule in [SEQ ID NO:1], except that Sequence A may include up to 10 point mutations, such as substitutions with other nucleotides, per each 100 nucleotides of the referenced portion of the nucleic acid molecule in [SEQ ID NO:1]. Mutations described in this application preferably do not disrupt the reading

15 frame of the coding sequence. Nucleic acid molecules functionally equivalent to the CSP sequences can occur in a variety of forms as described below.

Nucleic acid molecules may encode conservative amino acid changes in CSP peptide. The invention includes functionally equivalent nucleic acid molecules that encode conservative amino acid changes within a CSP amino acid sequence and

20 produce silent amino acid changes in CSP.

Nucleic acid molecules may encode non-conservative amino acid substitutions, additions or deletions in CSP peptide. The invention includes functionally equivalent nucleic acid molecules that make non-conservative amino acid changes within the CSP amino acid sequence in [SEQ ID NO:2]. Functionally equivalent nucleic acid molecules

25 include DNA and RNA that encode peptides, peptides and proteins having non-conservative amino acid substitutions (preferably substitution of a chemically similar amino acid), additions, or deletions but which also retain the same or similar CSP activity as the CSP peptide shown in [SEQ ID NO:2]. The DNA or RNA can encode fragments or variants of CSP. Fragments are useful as immunogens and in immunogenic

30 compositions (U.S. Patent No. 5,837,472). The CSP or CSP-like activity of such fragments and variants is identified by assays as described below. Fragments and variants of CSP encompassed by the present invention should preferably have at least about 40%, 60%, 80% or 95% sequence identity to the naturally occurring CSP nucleic

acid molecule, or a region of the sequence, such as the coding sequence or one of the conserved domains of the nucleic acid molecule, without being identical to the sequence in [SEQ ID NO:1]. Sequence identity is preferably measured with the ClustalW program (preferably using default parameters) (Thompson, JD et al., Nucleic Acid Res. 22:4673-4680)

Nucleic acid molecules functionally equivalent to the CSP nucleic acid molecule in [SEQ ID NO:1] will be apparent from the following description. For example, the sequence shown in [SEQ ID NO:1] may have its length altered by natural or artificial mutations such as partial nucleotide insertion or deletion, so that when the entire length of the coding sequence within [SEQ ID NO:1], is taken as 100%, the functional equivalent nucleic acid molecule preferably has a length of about 60-120% thereof, more preferably about 80-110% thereof. Fragments may be less than 60%.

Nucleic acid molecules containing partial (usually 80% or less, preferably 60% or less, more preferably 40% or less of the entire length) natural or artificial mutations so that some codons in these sequences code for different amino acids, but wherein the resulting peptide retains the same or similar CSP activity as that of a naturally occurring CSP peptide. The mutated DNAs created in this manner should preferably encode a peptide having at least about 40%, preferably at least about 60%, at least about 80%, and more preferably at least about 90% or 95% sequence identity to the amino acid sequence of the CSP peptide in [SEQ ID NO:2]. The ClustalW program preferably assesses sequence identity.

Since the genetic code is degenerate, the nucleic acid sequence in [SEQ ID NO:1] is not the only sequence which may code for a peptide having CSP activity. This invention includes nucleic acid molecules that have the same essential genetic information as the nucleic acid molecule described in [SEQ ID NO:1]. Nucleic acid molecules (including RNA) having one or more nucleic acid changes compared to the sequences described in this application and which result in production of a peptide shown in [SEQ ID NO:2] are within the scope of the invention.

Other functional equivalent forms of CSP-encoding nucleic acids can be isolated using conventional DNA-DNA or DNA-RNA hybridization techniques. Thus, the present invention also includes nucleic acid molecules that hybridize to one or more of the sequences in [SEQ ID NO:1] or its complementary sequence, and that encode expression for peptides, peptides and proteins exhibiting the same or similar activity as

that of the CSP peptide produced by the DNA in [SEQ ID NO:1] or its variants. Such nucleic acid molecules preferably hybridize to the sequence in [SEQ ID NO:1] under moderate to high stringency conditions (see Sambrook et al. Molecular Cloning: A Laboratory Manual, Most Recent Edition, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.). High stringency washes have low salt (preferably about 0.2% SSC), and low stringency washes have high salt (preferably about 2% SSC). A temperature of about 37 °C or about 42 °C is considered low stringency, and a temperature of about 50-65 °C is high stringency. The invention also includes a method of identifying nucleic acid molecules encoding a CSP activator peptide (preferably a mammalian peptide), including contacting a sample containing nucleic acid molecules including all or part of [SEQ ID NO:1] (preferably at least about 15 or 20 nucleotides of [SEQ ID NO:1]) under moderate or high stringency hybridization conditions and identifying nucleic acid molecules which hybridize to the nucleic acid molecules including all or part of [SEQ ID NO:1].). Similar methods are described in U.S. Patent No. 5,851,788, which is incorporated by reference in its entirety.

The invention also includes methods of using all or part of the nucleic acid molecules which hybridize to all or part of [SEQ ID NO:1], for example as probes or in assays to identify antagonists or inhibitors of the peptides produced by the nucleic acid molecules (described below). The invention also includes methods of using nucleic acid molecules having sequence identity to the CSP nucleic acid molecule (as described below) in similar methods.

The invention also includes a nucleic acid molecule detection kit including, preferably in a suitable container means or attached to a surface, a nucleic acid molecule of the invention encoding CSP or a peptide having CSP activity and a detection reagent (such as a detectable label). Other variants of kits will be apparent from this description and teachings in patents such as U.S. Patent Nos. 5,837,472 and 5,801,233, which are incorporated by reference in their entirety.

A nucleic acid molecule described above is considered to have a function substantially equivalent to the CSP nucleic acid molecules of the present invention if the peptide produced by the nucleic acid molecule has CSP activity. A peptide has CSP activity if it can stimulate genetic competence and acid tolerance in *S. mutans*. Activation of the HK/RR is shown where a peptide is capable of stimulating the uptake and incorporation of foreign DNA. We describe below how the activity of these peptide-mediated processes can be measured by determining the efficiency of plasmid uptake, which is a

measure of genetic competence. Since the ability to transport and incorporate foreign DNA relies on activation of the HK/RR and subsequent genes activated by the signal cascade initiated by the signal peptide, measurement of the conferment of erythromycin resistance by cells exposed to the peptide and plasmid DNA conferring erythromycin resistance indicates its level of function. Conversely if an inhibitor is capable of interfering with the action of the peptide the competence assay will indicate this by a corresponding decrease in the number of cells that acquire erythromycin resistance as described in the assays below (assays of genetic competence and assay of transformation of biofilms). Activation of the HK/RR is also shown where a peptide is capable of stimulating an acid tolerance response. We describe below how the activity of these peptide-mediated processes can be measured by determining the survival rate of cells in acidic pH conditions. Since the ability to survive exposure to acidic pH depends on the activation of the HK/RR and subsequent genes activated by the signal peptide, measurement of the survival of *S. mutans* in low pH conditions indicates the level of function of the signal peptide. Conversely, if an inhibitor is capable of interfering with the signal peptide sensing system the assay for acid adaptation will indicate this by a corresponding decrease in the survival rate of cells grown in acidic pH conditions as described in the assay below (assay of acid adaptation).

#### **Production of CSP in eukaryotic and prokaryotic cells**

The nucleic acid molecules of the invention may be obtained from a cDNA library. The nucleotide molecules can also be obtained from other sources known in the art such as expressed sequence tag analysis or *in vitro* synthesis. The DNA described in this application (including variants that are functional equivalents) can be introduced into and expressed in a variety of eukaryotic and prokaryotic host cells. A recombinant nucleic acid molecule for the CSP contains suitable operatively linked transcriptional or translational regulatory elements. Suitable regulatory elements are derived from a variety of sources, and they may be readily selected by one with ordinary skill in the art (Sambrook, J, Fritsch, E.E. & Maniatis, T. (Most Recent Edition). Molecular Cloning: A Laboratory Manual. Cold Spring Harbor Laboratory Press. New York; Ausubel et al. (Most Recent Edition). Current Protocols in Molecular Biology, John Wiley & Sons, Inc.). For example, if one were to upregulate the expression of the nucleic acid molecule, one could insert a sense sequence and the appropriate promoter into the vector. Promoters can be inducible or constitutive, environmentally- or developmentally-regulated, or cell- or tissue-specific. Transcription is enhanced with promoters known in the art for

expression. The CMV and SV40 promoters are commonly used to express desired peptide in cells. Other promoters known in the art may also be used (many suitable promoters and vectors are described in the applications and patents referenced in this application).

5           If one were to downregulate the expression of the nucleic acid molecule, one could insert the antisense sequence and the appropriate promoter into the vehicle. The nucleic acid molecule may be either isolated from a native source (in sense or antisense orientations), synthesized, or it may be a mutated native or synthetic sequence or a combination of these.

10           Examples of regulatory elements include a transcriptional promoter and enhancer or RNA polymerase binding sequence, a ribosomal binding sequence, including a translation initiation signal. Additionally, depending on the vector employed, other genetic elements, such as selectable markers, may be incorporated into the recombinant molecule. Other regulatory regions that may be used include an enhancer domain and a  
15           termination region. The regulatory elements may be bacterial, fungal, viral or avian in origin. Likewise the regulatory elements may originate from animal, plant, yeast, insect or other sources, including synthetically produced elements and mutated elements.

          In addition to using the expression vectors described above, the peptide may be expressed by inserting a recombinant nucleic acid molecule in a known expression system  
20           derived from bacteria, viruses, yeast, mammals, insects, fungi or birds. The recombinant molecule may be introduced into the cells by techniques such as *Agrobacterium tumefaciens*-mediated transformation, particle-bombardment-mediated transformation, direct uptake, microinjection, coprecipitation, transfection and electroporation depending on the cell type. Retroviral vectors, adenoviral vectors, Adeno Associated Virus (AAV)  
25           vectors, DNA virus vectors and liposomes may be used. Suitable constructs are inserted in an expression vector, which may also include markers for selection of transformed cells. The construct may be inserted at a site created by restriction enzymes.

          In one embodiment of the invention, a cell is transfected with a nucleic acid molecule of the invention inserted in an expression vector to produce cells expressing a  
30           peptide encoded by the nucleic acid molecule.

          Another embodiment of the invention relates to a method of transfecting a cell with a nucleic acid molecule of the invention, inserted in an expression vector to produce

a cell expressing the CSP peptide or other peptide of the invention. The invention also relates to a method of expressing the peptides of the invention in a cell. A preferred process would include culturing a cell including a recombinant DNA vector including a nucleic acid molecule encoding CSP (or another nucleic acid molecule of the invention) in a culture medium so that the peptide is expressed. The process preferably further includes recovering the peptide from the cells or culture medium.

### Probes

The invention also includes oligonucleotide probes made from the cloned CSP nucleic acid molecules described in this application or other nucleic acid molecules of the invention (see Materials and Methods section). The probes may be 15 to 20 nucleotides in length. A preferred probe is at least 15 nucleotides of CSP in [SEQ ID NO:1]. The invention also includes at least 15 consecutive nucleotides of [SEQ ID NO:1]. The probes are useful to identify nucleic acids encoding CSP peptides as well as peptides functionally equivalent to CSP. The oligonucleotide probes are capable of hybridizing to the sequence shown in [SEQ ID NO:1] under stringent hybridization conditions. A nucleic acid molecule encoding a peptide of the invention may be isolated from other organisms by screening a library under moderate to high stringency hybridization conditions with a labeled probe. The activity of the peptide encoded by the nucleic acid molecule is assessed by cloning and expression of the DNA. After the expression product is isolated, the peptide is assayed for CSP activity as described in this application.

Functionally equivalent CSP nucleic acid molecules from other cells, or equivalent CSP-encoding cDNAs or synthetic DNAs, can also be isolated by amplification using Polymerase Chain Reaction (PCR) methods. Oligonucleotide primers, such as degenerate primers, based on [SEQ ID NO:1] can be prepared and used with PCR and reverse transcriptase (E. S. Kawasaki (1990), In Innis et al., Eds., PCR Protocols, Academic Press, San Diego, Chapter 3, p. 21) to amplify functional equivalent DNAs from genomic or cDNA libraries of other organisms. The oligonucleotides can also be used as probes to screen cDNA libraries.

### Functionally equivalent peptides, peptides and proteins

The present invention includes not only the peptides encoded by the sequences of the invention, but also functionally equivalent peptides, peptides and proteins that exhibit the same or similar CSP peptide activity. A peptide is considered to possess a function substantially equivalent to that of the CSP peptide if it has CSP activity. CSP activity means that it is able to confer genetic competence to *S. mutans*, as measured by an increased ability to incorporate and express foreign genetic material, when added to cells as described in the assay of genetic competence below. CSP activity also means that the peptide is able to confer an acid tolerance response in *S. mutans* as measured by an increase in cell survival under acidic pH conditions when added to cells as described in the assay for acid adaptation below. Functionally equivalent peptides, peptides and proteins include peptides, peptides and proteins that have the same or similar protein activity as CSP when assayed, i.e. they are able to stimulate genetic competence and low pH tolerance (the ability to withstand acid challenges of pH 3.5 – pH 3.0 for up to 3 hours) in *S. mutans*. A peptide has CSP activity if it is capable of increasing the frequency of uptake and expression of foreign DNA as described in the following assay for genetic competence and if the peptide can promote an acid tolerance response as described in the assay for acid adaptation.

Identity refers to the similarity of two peptides or proteins that are aligned so that the highest order match is obtained. Identity is calculated according to methods known in the art, such as the ClustalW program. For example, if a peptide (called "Sequence A") has 90% identity to a portion of the peptide in [SEQ ID NO:2], then Sequence A will be identical to the referenced portion of the peptide in [SEQ ID NO:2], except that Sequence A may include up to 1 point mutations, such as substitutions with other amino acids, per each 10 amino acids of the referenced portion of the peptide in [SEQ ID NO:2]. Peptides, peptides and proteins functionally equivalent to the CSP peptides can occur in a variety of forms as described below.

Peptides biologically equivalent in function to CSP peptide include amino acid sequences containing amino acid changes in the CSP sequence. The functionally equivalent peptides have at least about 40% sequence identity, preferably at least about 60%, at least about 75%, at least about 80%, at least about 90% or at least about 95% sequence identity, to the natural CSP peptide or a corresponding region. The ClustalW



program preferably determines sequence identity. Most preferably, 1, 2, 3, 4, 5, 5-10, 10-15 amino acids are modified.

Variants of the CSP peptide may also be created by splicing. A combination of techniques known in the art may be used to substitute, delete or add amino acids. For

5 example, a hydrophobic residue such as methionine can be substituted for another hydrophobic residue such as alanine. An alanine residue may be substituted with a more hydrophobic residue such as leucine, valine or isoleucine. An aromatic residue such as phenylalanine may be substituted for tyrosine. An acidic, negatively-charged amino acid such as aspartic acid may be substituted for glutamic acid. A positively-  
10 charged amino acid such as lysine may be substituted for another positively-charged amino acid such as arginine. Modifications of the peptides of the invention may also be made by treating a peptide of the invention with an agent that chemically alters a side group, for example, by converting a hydrogen group to another group such as a hydroxy or amino group.

15 Peptides having one or more D-amino acids are contemplated within the invention. Also contemplated are peptides where one or more amino acids are acetylated at the N-terminus. Those skilled in the art recognize that a variety of techniques are available for constructing peptide mimetics (i.e., a modified peptide or peptide or protein) with the same or similar desired biological activity as the  
20 corresponding peptide of the invention but with more favorable activity than the peptide with respect to characteristics such as solubility, stability, and/or susceptibility to hydrolysis and proteolysis. See for example, Morgan and Gainor, *Ann. Rep. Med. Chem.*, 24:243-252 (1989).

The invention also includes hybrid nucleic acid molecules and peptides, for  
25 example where a nucleic acid molecule from the nucleic acid molecule of the invention is combined with another nucleic acid molecule to produce a nucleic acid molecule which expresses a fusion peptide. One or more of the other domains of CSP described in this application could also be used to make fusion peptides. For example, a nucleotide domain from a molecule of interest may be ligated to all or part of a nucleic acid  
30 molecule encoding CSP peptide (or a molecule having sequence identity) described in this application. Fusion nucleic acid molecules and peptides can also be chemically synthesized or produced using other known techniques. The invention includes a

nucleic acid molecule encoding a fusion peptide or a recombinant vector including the nucleic acid molecule

The variants preferably retain the same or similar CSP activity as the naturally occurring CSP. The CSP activity of such variants can be assayed by techniques  
5 described in this application and known in the art.

Variants produced by combinations of the techniques described above but which retain the same or similar CSP activity as naturally occurring CSP are also included in the invention (for example, combinations of amino acid additions, and substitutions).

Variants of CSP encompassed by the present invention preferably have at least  
10 about 40% sequence identity, preferably at least about 60%, 75%, 80%, 90% or 95% sequence identity, to the naturally occurring peptide, or corresponding region or moiety of the peptide, or corresponding region. Sequence identity is preferably measured with the ClustalW.

#### 15 **Histidine Kinase & Response Regulator**

The invention also includes sequences having identity with the histidine kinase, response regulator of the invention and comA and comB. Preferred percentages of identity (nucleic acid molecule and polypeptide) are the same as those described for the CSP.

As well, probes and antibodies for a histidine kinase, response regulator comA or comB may be prepared using the description in this application and techniques known in the art. The description for preparation of CSP variants and mutants is also applicable to the histidine kinase, response regulator or comA and comB of the invention. The invention also includes fragments of HK having HK activity, fragments of RR having RR  
20 activity and fragments of comA or comB having activity.  
25

#### **Design of CSP peptide competitive inhibitors**

The activity of the CSP peptide may be varied by carrying out selective site-directed mutagenesis. We characterize the binding domain and other critical amino  
30 acid residues in the peptide that are candidates for mutation, insertion and/or deletion. Sequence variants may be synthesized. A DNA plasmid or expression

vector containing the CSP nucleic acid molecule or a nucleic acid molecule having sequence identity may be used for these studies using the U.S.E. (Unique site elimination) mutagenesis kit from Pharmacia Biotech or other mutagenesis kits that are commercially available, or using PCR. Once the mutation is created and confirmed by DNA sequence analysis, the mutant peptide is expressed using an expression system and its activity is monitored. This approach is useful to identify CSP inhibitors. All these modifications of the CSP DNA sequences presented in this application and the peptides produced by the modified sequences are encompassed by the present invention.

10

#### **Pharmaceutical compositions**

The CSP inhibitors are also useful when combined with a carrier in a pharmaceutical composition. The compositions are useful when administered in methods of medical treatment or prophylaxis of a disease, disorder or abnormal physical state caused by *S. mutans*. The invention also includes methods of medical treatment of a disease, disorder or abnormal physical state characterized by excessive *S. mutans* or levels or activity of CSP peptide, for example by administering a pharmaceutical composition including a carrier and a CSP inhibitor. Caries is one example of a disease, which can be treated or prevented by antagonizing CSP.

20

The pharmaceutical compositions can be administered to humans or animals by methods such as food, food additives, gel, toothpaste, mouthwash, dental floss or chewing gum in methods of medical treatment. The peptides of the invention may be coupled to lipids or carbohydrates. This increases their ability to adhere to teeth, either by prolonging the duration of the adhesion or by increasing its affinity, or both. They may also be coupled to polymers, for example in dental work (eg. crowns, braces, fillings) or dental floss. The pharmaceutical compositions can be administered to humans or animals. Dosages to be administered depend on individual patient condition, indication of the drug, physical and chemical stability of the drug, toxicity of the desired effect and the chosen route of administration (Robert Rakel, ed., Conn's Current Therapy (1995, W.B. Saunders Company, USA)). The pharmaceutical compositions are used to treat diseases caused by streptococcal infections such as caries and endocarditis.

30

CSP activity could be blocked by antisense mRNA or by inhibiting the activity of the exporter that secretes it from the cell. We have the sequence of these exporters. There are two copies of the genes (comAB) that are involved in export.

Nucleic acid molecules (antisense inhibitors of CSP) and competitive inhibitors of CSP may be introduced into cells using *in vivo* delivery vehicles such as liposomes. They may also be introduced into these cells using physical techniques such as microinjection and electroporation or chemical methods such as coprecipitation or using liposomes.

The pharmaceutical compositions can be prepared by known methods for the preparation of pharmaceutically acceptable compositions which can be administered to patients, and such that an effective quantity of the nucleic acid molecule or peptide is combined in a mixture with a pharmaceutically acceptable vehicle. Suitable carriers are described, for example in Remington's Pharmaceutical Sciences (Remington's Pharmaceutical Sciences, Mack Publishing Company, Easton, Pa., USA). Carriers include saline and D5W (5% dextrose and water). Excipients include additives such as a buffer, solubilizer, suspending agent, emulsifying agent, viscosity controlling agent, flavor, lactose filler, antioxidant, preservative or dye. There are preferred excipients for stabilizing peptides for parenteral and other administration. The excipients include serum albumin, glutamic or aspartic acid, phospholipids and fatty acids.

On this basis, the pharmaceutical compositions could include an active compound or substance, such as a CSP inhibitor, in association with one or more pharmaceutically acceptable vehicles or diluents, and contained in buffered solutions with a suitable pH and isoosmotic with the physiological fluids. The methods of combining the active molecules with the vehicles or combining them with diluents is well known to those skilled in the art. The compositions may also contain additives such as antioxidants, buffers, bacteriostats, bactericidal antibiotics and solutes which render the formulation isotonic in the intended recipient; and aqueous and non-aqueous sterile suspensions which may include suspending agents and thickening agents. The composition could include a targeting agent for the transport of the active compound to specified sites.

### 30 Vaccines

Antibodies directed against the CSP would provide protection against caries. Antibodies may be manufactured as described below. Alternatively, a peptide of the

invention or a fragment thereof may be used with a carrier to make a vaccine. The peptide or fragment may also be conjugated to another molecule to increase its antigenicity. Antibodies can also be coupled to the peptide (Brady, L.J. et al., "Monoclonal Antibody-Mediated Modulation of the Humoral Immune Response against Mucosally Applied *Streptococcus mutans*" (in press). In order to enhance the immune response the peptide can be coupled to KLH, ovalbumin, or thyroglobulin prior to immunization. The vaccine composition will trigger the mammal's immune system to produce antibodies. The invention includes vaccine compositions and methods of vaccinating a mammal, preferably a human, against dental caries by administering to the mammal an effective amount of a vaccine composition. Techniques for preparing and using vaccines are known in the art. To prepare the vaccine, the peptide, or a fragment of the peptide, may be mixed with other antigens (of different immunogenicity), a vehicle or an excipient. Examples of peptide vaccines are found in U.S. Patent Nos. 5,679,352, 5,194,254 and 4,950,480. Techniques for preparing vaccines involving site-directed mutagenesis are described in U.S. Patent Nos. 5,714,372, 5,543,302, 5,433,945, 5,358,868, 5,332,583, 5,244,657, 5,221,618, 5,147,643, 5,085,862 and 5,073,494. Vaccines may be administered by known techniques, such as topical or parenteral administration. Vast changes are taking place in vaccinology consequent to the introduction of new technologies. Acellular purified fractions devoid of side effects, non-pathogenic but immunogenic mutants, recombinant technology, conjugated vaccines, combination vaccines (to limit the number of injections). Vaccine delivery systems can deliver multiple doses of the vaccine at a single contact point. A genetically engineered oral vaccine is useful to impart better and longer duration of immunity. Oral vaccines are useful. The nose as a route for immunization is also useful. DNA alone can constitute the vaccines, inducing both humoral and cell-mediated immune responses. Live recombinant vaccines are also useful. Potent adjuvants add to the efficacy of the vaccines. One can also 'humanize' mouse monoclonals by genetic engineering and express these efficiently in plants. These recombinant antibodies are opening out an era of highly specific and safe therapeutic interventions. An advantage of preformed antibodies directed at a defined target and given in adequate amounts is the certainty of efficacy in every recipient, in contrast to vaccines, where the quality and quantum of immune response varies from individual to individual. For example, nasal immunization may be done as described in C. Jespersgaard et al. "Protective Immunity against *Streptococcus mutans* Infection in Mice after Intranasal Immunization with the Glucan-Binding Region of *S. mutans*

Glucosyltransferase" Infection and Immunity, December 1999, p. 6543-6549, Vol. 67, No.

12. Vaccine compositions may comprise solid or liquid formulations such as gels, sprays, inhalants, tablets, toothpastes, mouthwashes or chewing gum.

For vaccine application, cholera toxin can be used by coupling the peptide to its

5 B-subunit to stimulate production of secretory antibody i.e., Coupling to CTB.

### Screening for inhibitors of CSP

Inhibitors are preferably directed towards CSP to block *S. mutans* competence, low pH tolerance and biofilm formation.

- 10 A method of identifying a compound which reduces the interaction of CSP with HK, can include: contacting (i) CSP with (ii) HK, a CSP-binding fragment of HK or a derivative of either of the foregoing in the presence of the compound; and b) determining whether the interaction between (i) and (ii) is reduced, thereby indicating that the compound reduces the interaction of CSP and HK. A CSP inhibitor
- 15 (carries treating or preventing compound) inhibits the interaction between (i) and (ii). By way of example, one can screen a synthetic peptide library. One could also screen small non-peptide organic molecules.

- In one embodiment, the invention includes an assay for evaluating whether test compounds are capable of acting as agonists or antagonists for CSP, or a peptide
- 20 having CSP functional activity, including culturing cells containing DNA which expresses CSP, or a peptide having CSP activity so that the culturing is carried out in the presence of at least one compound whose ability to modulate CSP activity is sought to be determined and thereafter monitoring the cells for either an increase or decrease in the level of CSP or CSP activity. Other assays (as well as variations of the above assay)
- 25 will be apparent from the description of this invention and techniques such as those disclosed in U.S. Patent No. 5,851,788, 5,736,337 and 5,767,075 which are incorporated by reference in their entirety. For example, the test compound levels may be either fixed or variable.

### Preparation of antibodies

The CSP peptide is also useful as an antigen for the preparation of antibodies that can be used to purify or detect other CSP-like peptides. Antibodies may also block CSP binding to HK. Antibodies are preferably targeted to the entire CSP sequence.

- 5 The CSP peptide may be conjugated to other compounds, in order to increase immunogenicity.

We generate polyclonal antibodies against the CSP, which is a unique sequence. Monoclonal and polyclonal antibodies are prepared according to the description in this application and techniques known in the art. For examples of methods of preparation  
10 and uses of monoclonal antibodies, see U.S. Patent Nos. 5,688,681, 5,688,657, 5,683,693, 5,667,781, 5,665,356, 5,591,628, 5,510,241, 5,503,987, 5,501,988, 5,500,345 and 5,496,705, which are incorporated by reference in their entirety. Examples of the preparation and uses of polyclonal antibodies are disclosed in U.S. Patent Nos. 5,512,282, 4,828,985, 5,225,331 and 5,124,147 which are incorporated by  
15 reference in their entirety. Antibodies recognizing CSP can be employed to screen organisms or tissues containing CSP peptide or CSP-like peptides. The antibodies are also valuable for immuno-purification of CSP or CSP-like peptides from crude extracts.

An antibody (preferably the antibody described above) may be used to detect CSP or a similar peptide, for example, by contacting a biological sample with the  
20 antibody under conditions allowing the formation of an immunological complex between the antibody and a peptide recognized by the antibody and detecting the presence or absence of the immunological complex whereby the presence of CSP or a similar peptide is detected in the sample. The invention also includes compositions preferably including the antibody, a medium suitable for the formation of an immunological complex  
25 between the antibody and a peptide recognized by the antibody and a reagent capable of detecting the immunological complex to ascertain the presence of CSP or a similar peptide. The invention also includes a kit for the *in vitro* detection of the presence or absence of CSP or a similar peptide in a biological sample, wherein the kit preferably includes an antibody, a medium suitable for the formation of an immunological complex  
30 between the antibody and a peptide recognized by the antibody and a reagent capable of detecting the immunological complex to ascertain the presence of CSP or a similar peptide in a biological sample. Further background on the use of antibodies is provided,

for example in U.S. Patent Nos. 5,695,931 and 5,837,472, which are incorporated by reference in their entirety.

#### Research Tool

##### 5 Assay of genetic competence

The ability of the peptide to activate the HK and RR and the subsequent genes involved in the conferral of the properties of genetic competence, acid tolerance and biofilm formation can be determined by measuring the efficiency of uptake and expression of DNA (preferably plasmid DNA) in *S. mutans* when exposed to signal peptide and/or  
10 inhibitor. Two methods modified based on the protocols described by Perry *et al.* Infect Immun, 41:722-727 and Lindler and Macrina J Bacteriol, 166:658-665 are used to assay genetic competence. The method involves adding DNA and CSP (preferably plasmid DNA) to a *S. mutans* culture (or culture of a bacteria expressing CSP or a variant thereof). The rate of transformation is then determined. *S. mutans* is preferably grown in  
15 THYE plus 5% horse serum (THYE-HS). After 2-hr incubation, 1 µg/ml plasmid DNA or 10 µg/ml of chromosomal DNA is added to the culture. To assay induction of competence, competence signal peptide, (SCSP) is then added to the cultures, incubation continued for 30 minutes with a final concentration of 500 ng/ml of SCSP added to each sample. After the 30-minute incubation equal amounts of DNA is added to  
20 each well (1 µg/ml plasmid or 10 µg/ml of chromosomal DNA) and incubation continued for another 2 hrs.. Cell dilutions were immediately spread on THYE agar plates plus appropriate antibiotics. Transformation frequency was expressed as the number of transformants (antibiotic resistant cells) per number of viable recipients. This is determined by comparing the number of cells able to grow in the presence of antibiotic  
25 (conferred by the applied plasmid or chromosomal DNA) relative to the total number of cells present (i.e., that grow in the absence of antibiotic). A higher value indicates a higher rate of transformation and thus is reflective of a stimulatory effect by the peptide. Consequently, addition of a molecule that successfully acts as an inhibitor results in a lower ratio of transformants/recipients, indicating that the inhibitor is effective at blocking  
30 activity of the CSP. CSP deficient cells may also be used in a variation of these assays. One can identify compounds that inhibit CSP or variants thereof by adding a test compound to the mixture to determine if the rate of transformation is decreased by the addition of the test compound.



The activity of the system can also be measured by an *in vitro* assay that relies on the measurement of marker protein expression (such as green fluorescent protein (GFP)) via expression from a fusion to a promoter controlled by the signal cascade initiated by CSP/HK/RR. One such promoter occurs immediately 5' proximal to the *S.*

5 *mutans* comX gene. *S. mutans* cells grown in microtiter wells are exposed to the CSP and/or inhibitor and the level of fluorescence of the comX::GFP strain is measured to give a quantitative measure of CSP stimulation (and conversely inhibitor activity). One can identify compounds that inhibit CSP or variants thereof by adding a test compound to the mixture to determine if the quantitative measure of CSP stimulation is decreased  
10 by the addition of the test compound.

#### *Assay of acid resistance tolerance*

The ability of CSP to promote acid resistance tolerance is determined by measuring the cell survival rate of *S. mutans* when exposed to acidic pH. In one  
15 example, *S. mutans* are first grown in batch culture to assay acid tolerance response in 'standard' log- and stationary-phase cells by using a modification of methods described previously by Svensäter et al. Oral Microbiol. Immunol., 12:266-73. Mid-log-phase cells are obtained by transferring one volume of overnight culture into nine volumes (1:10) of fresh TYG medium (pH 7.5) and incubated at 37°C with 5% CO<sub>2</sub> for 2 hours. These  
20 cells are then collected by centrifugation at 8,000 x g for 10 min and resuspended in 2 ml of fresh TYG (pH 5.5) at various cell densities as determined by O.D<sub>600</sub>. The cells are induced for acid adaptation by incubation at pH 5.5 for 2 h at 37°C with 5% CO<sub>2</sub>. The adapted log-phase cells are then exposed to the killing pH. Killing pH is pre-determined by incubating unadapted, mid-log phase cells in TYG medium at pH values from 6.0 to  
25 2.0. Stationary-phase cells are prepared by re-suspending late-log phase cells in TY medium (tryptone-yeast extract) without glucose. The culture is incubated at 37°C for 2 h to allow the cells to fully enter into stationary phase. Induction of acid adaptation in stationary-phase cells follows a similar procedure to that for log-phase cells. Adaptation of both log- and stationary-phase cells to acidic pH is determined by measuring the  
30 ability of bacterial cells to survive a killing pH for 3 h. Acid killing is initiated by resuspending cells in the same volume of fresh TYG (pH 3.5) and an aliquot of cell suspension is taken immediately from each sample to determine total viable cell number at zero time. The cells are then incubated for 3 h at 37°C with 5% CO<sub>2</sub> and an aliquot of sample is taken to determine survival rate by viable cell counts. Addition of a molecule

that successfully acts as an inhibitor results in a decrease in the acid resistance tolerance of *S mutans* resulting in a corresponding decrease in cell survival indicating that the inhibitor is effective at blocking activity of CSP. CSP deficient cells may also be used in a variation of these assays wherein addition of the signal peptide can  
5 complement the acid-adaptation-defective phenotype of a *comC* deficient cell. One can identify compounds that inhibit CSP or variants thereof by adding a test compound to the mixture to determine if the survival rate of cells is decreased by the addition of the test compound

Cells transformed with a nucleic acid molecule of the invention (histidine kinase,  
10 CSP or response regulator) are useful as research tools. For example, one may obtain a cell (or a cell line, such as an immortalized cell culture or a primary cell culture) that does not express histidine kinase, CSP or response regulator, insert a histidine kinase, CSP or response regulator nucleic acid molecule in the cell, and assess the level of expression and activity. Alternatively, histidine kinase, CSP or response regulator  
15 nucleic acid molecules may be over-expressed in a cell that expresses a histidine kinase, CSP or response regulator nucleic acid molecule. In another example, experimental groups of cells may be transformed with vectors containing different types of histidine kinase, CSP or response regulator nucleic acid molecules to assess the levels of polypeptides and peptides produced, its functionality and the phenotype of the  
20 cells. The polypeptides and peptides are also useful for *in vitro* analysis of histidine kinase, CSP or response regulator activity or structure. For example, the polypeptides and peptides produced can be used for microscopy or X-ray crystallography studies.

The histidine kinase, CSP or response regulator nucleic acid molecules and polypeptides are also useful in assays for the identification and development of  
25 compounds to inhibit and/or enhance polypeptide or peptide function directly. For example, they are useful in an assay for evaluating whether test compounds are capable of acting as antagonists for histidine kinase, CSP or response regulator by: (a) culturing cells containing a nucleic acid molecule which expresses histidine kinase, CSP or response regulator peptides (or fragments or variants thereof having histidine kinase,  
30 CSP or response regulator activity) wherein the culturing is carried out in the presence of increasing concentrations of at least one test compound whose ability to inhibit histidine kinase, CSP or response regulator is sought to be determined; and (b) monitoring in the cells the level of inhibition as a function of the concentration of the test compound,

thereby indicating the ability of the test compound to inhibit histidine kinase, CSP or response regulator activity.

Suitable assays may be adapted from, for example, US patent no. 5,851,788.

## 5 MATERIALS AND METHODS:

### *Growth conditions of cells*

Cells are grown in Todd Hewitt yeast extract medium at various dilutions with and without 5% horse serum and 0.01% hog gastric mucin.

10

### *Protocol for Transformation of Biofilm-grown Cells*

Biofilms are developed on polystyrene microtiter plates to provide a rapid and simple method for assaying biofilm formation, and hence activity of the peptide/receptor/kinase system. Formation of biofilms is initiated by inoculating 20  $\mu$ l of cell suspension into each well containing 2 ml of biofilm medium (4X diluted Todd-Hewitt Yeast Extract supplemented with final concentration of 0.01% hog gastric mucin) for overnight incubation at 37°C under an anaerobic condition. After 20-h incubation, fluid medium is removed and added with 2 ml of pre-warmed, fresh THYE plus 5% horse serum. The cultures are incubated for 30 minutes and each well is supplemented with a final concentration of 200 ng/ml of synthetic competence stimulating peptide (SCSP) and varying concentrations of the inhibitor and the incubation is continued. After 30 minutes, plasmid DNA (1 mg/ml) or chromosomal DNA (10 mg/ml) is added to each well and the cultures are incubated for an additional 2 hr. Planktonic cells are then removed and the wells are washed once with PBS buffer. Biofilm cells are collected into 2 ml fresh medium by a gentle sonication or washing the wells using a pipette. The samples are centrifuged at 12,000 x g for 5 min. Both biofilm and planktonic cells are resuspended into 200  $\mu$ l of fresh medium and are immediately spread on THYE agar plus appropriate antibiotics. Transformation frequency is determined after 48-h of incubation.

25

### *Genome database analysis*

30 Homologues of the *Streptococcus pneumoniae* comD/E genes encoding a histidine kinase/ response regulator system were identified. This sequence was used to design primers to amplify the region from a number of *S. mutans* isolates. An open reading

frame consisting of 138 nucleotides was located 148 nucleotides 5' proximal from the end of the *comD* homolog in the opposite orientation (Fig 1). This ORF was found to encode a peptide of 46-amino acid in length, the precursor of the 21-amino acid CSP.

#### 5 *PCR amplification and nucleotide sequencing*

The *comCDE* genes were amplified from the genomes of several *S. mutans* isolates by PCR using primers designed based on the genome database sequence and their nucleotide sequences determined. The deduced amino acid sequences are compared among the isolates by sequence alignment to confirm identity.

10

#### *Gene inactivations*

Genes are inactivated by integration of internal homologous fragments into the suicide vector pVA8912. Mutants defective in each of the individual genes (*comC*, *comD*, *comE*) are inactivated and their phenotypes are compared to the parent strain NG8 for their abilities to form biofilms, tolerate acidic pH (pH 2-4), and transport and incorporate DNA. The knockout mutants of *comD* and *E* were constructed by insertion-duplication mutagenesis, whereas the knockout *comC* mutant was created by allelic exchange via insertion of an erythromycin resistance determinant into the *comC* locus (Li *et al*, 2001).

15

All mutant strains were therefore resistant to erythromycin. The wild-type strain was subcultured routinely on Todd-Hewitt-Yeast Extract (THYE) agar plates (BBL<sup>®</sup>; Becton Dickinson, Cockeysville, MD), whereas the mutants were maintained on THYE agar plus 10 µg/ml of erythromycin. A minimal medium (DMM) was prepared to grow biofilms by a modification of the method described previously (Loo *et al*, 2000). The medium contained 58 mM K<sub>2</sub>HPO<sub>4</sub>, 15 mM KH<sub>2</sub>PO<sub>4</sub>, 10 mM (NH<sub>4</sub>)<sub>2</sub>SO<sub>4</sub>, 35 mM NaCl, 2 mM MgSO<sub>4</sub>·7H<sub>2</sub>O, 0.2% (wt/vol) Casamino Acids and was supplemented with filter-sterilized vitamins, (0.04 mM nicotinic acid, 0.1 mM pyridoxine HCl, 0.01 mM pantothenic acid, 1 µM riboflavin, 0.3 µM thiamin HCl, and 0.05 µM D-biotin), amino acids (4 mM L-glutamic acid, 1 mM L-arginine HCl, 1.3 mM L-cysteine HCl, and 0.1 mM L-tryptophan) and 20 mM glucose.

20

25

30

#### *Synthesis of synthetic peptide*

The sequence of the processed peptide was deduced by determining the cleavage site to be located beside the gly-gly amino acid residues (numbers 24 and 25) (fig. 4). A

peptide was synthesized corresponding to amino acid sequence of residues 26-46 inclusive.

#### ***Restoration of phenotypic defects by addition of CSP***

- 5 To determine if the synthetic peptide could restore defective phenotypes of the *comC* mutants, a chemically synthesized 21-amino acid competence-stimulating peptide (CSP) (Li et al, 2001) was used in complementary experiments. The peptide was freshly dissolved in sterile distilled water to a concentration of 1 mg/ml. The CSP solution was then added to the cultures at a final concentration of 2 µg/ml 2 h after inoculation of  
10 bacterial cells.

#### ***Growth rates***

- The parent and mutant strains were grown in THYE medium for assaying their growth curves using a Bioscreen Microbiology Reader incorporating a multi-well disposable microtiter plate (Bioscreen C, Helsinki, Finland). The Bioscreen Reader was equipped  
15 with Biolink software program that allowed us to record and display the growth curves and growth rate calculations automatically. The growth of the strains was initiated by inoculating 5 µl of cell suspension into each well containing 200 µl of fresh THYE medium. The cell suspensions were pre-adjusted to the same optical density at O.D<sub>600</sub>  
20 before inoculation. The plates were then placed in the Bioscreen system, which was set up to read optical density automatically every 15 minutes with shaking. The readings of optical density were automatically recorded and converted into growth curves. Each assay was performed in quadruplicate.

#### ***Bacterial strains and growth conditions***

- Seven strains of *S. mutans* were used in this study (strains include: BM71, GB14, H7, JH1005, LT11, NG8, and UAB159. All the strains were cultured from freeze-dried ampoules and routinely maintained on Todd-Hewitt Yeast Extract (THYE) plates. For selection of antibiotic resistant colonies following transformation, the medium was  
30 supplemented with either erythromycin (Em) (10 µg/ml) or kanamycin (Km) (500 µg/ml).

#### ***Assay for biofilms formed on polystyrene microtiter plates***

Biofilms were developed on polystyrene microtiter plates to provide a rapid and simple method for assaying genetic transformation. A 4X diluted THYE medium supplied

with final concentration of 0.01% hog gastric mucin was used as biofilm medium (BM). Formation of biofilms was initiated by inoculating 20  $\mu$ l of cell suspension into each well containing 2 ml of BM and four wells were set up: two for assaying transformation and two for quantification of biofilms. After cultures were incubated at 37°C for 20 h under an anaerobic condition, fluid medium was removed for viable cell counts. The wells were rinsed once with 10mM PBS buffer (pH 7.2) and biofilm cells were collected in 2 ml PBS by a gentle sonication for 15 seconds. Both biofilm and the planktonic cells were immediately spread on THYE plates using a spiral system (Spiral Plater, Model D, Cincinnati, OH) and incubated at 37°C under an anaerobic condition. Formation of biofilms was quantified by viable cell counts after 48 h of incubation.

#### *Assay for "steady-state" biofilms*

Biofilms were also grown in a chemostat-based biofilm fermentor to define and optimize the conditions for genetic competence of biofilm-grown cells. The biofilm fermentor was modified in the Mechanical Engineering and Glass Blowing Shops, University of Toronto, based on a similar system described previously (Li and Bowden, 1994). The vessel was made of glass with a working volume of 400 ml. The vessel lip was constructed of stainless steel with 10 sampling ports, which allowed sterile insertion and retrieval of glass rods (0.5 cm in diameter, approximately 4.0 cm<sup>2</sup> area immersed in fluid medium), providing abiotic surfaces for accumulation of biofilms. Temperature in the chemostat vessel was maintained at 37°C  $\pm$  0.1 by a temperature controller (Model R-600F, Cole Parmer Instrument Cop., Vernon Hill, IL). The culture pH was controlled by a pH control unit (Digital pH Meter/Controller, Model 501-3400, Barnant Corp. Barrington, IL) through the addition of 1M KOH or 1M HCl. The vessel was placed on a magnetic stirrer (Fisher Scientific) and the culture was stirred at 200 rpm by a polypropylene coated magnetic stirrer bar (3 cm in length). Continuous cultures were obtained by pumping fresh 4X diluted THYE medium supplemented with a final concentration of 0.01% hog gastric mucin (Type III, Sigma) into the vessel (400 ml) at the desired dilution rates. Daily maintenance of the chemostat included optical density reading, viable cell counts and pH measurement in fluid cultures. When the cultures reached "steady-state" (at least 10 mean generation times), glass rods were aseptically inserted into the chemostat for the initiation of biofilm formation. Then, biofilms of different ages were removed from the cultures for both genetic transformation and quantification of biofilms using viable cell counts.

### **Scanning electron microscopy (SEM)**

To examine spatial distribution and biofilm thickness by scanning electron microscopy, 5  
biofilms of different ages were removed by slicing off the bottom of the microtiter wells  
that were then washed once with 10 mM KPO<sub>4</sub> and fixed with 2 ml of 3.7% formaldehyde  
in 10 mM KPO<sub>4</sub> buffer overnight. The samples were then dehydrated with a series of  
alcohol baths (30%, 50%, 70%, 95% and 100%), critical point dried with liquid CO<sub>2</sub>,  
mounted and sputter coated with gold. The samples were then examined using a  
10 scanning electron microscope (Model S-2500, Hitachi Instruments, San Jose, CA).

### **Transformation protocol**

Two methods modified based on the protocols described by Perry *et al* (Infect Immun,  
41:722-727) and Lindler and Macrina (J Bacteriol, 166:658-665) were used to assay  
15 natural transformation of biofilm cells. Biofilms formed on polystyrene microtiter plates  
were added with 2 ml of pre-warmed, fresh THYE plus 5% horse serum (THYE-HS)  
immediately following removal of the BM medium, and the incubation continued at 37°C.  
After 2h incubation, a final concentration of 1 µg/ml plasmid DNA or 10 µg/ml of  
chromosomal DNA was added to each well. The cultures were incubated for an  
20 additional 2 h before collection of the cells for plating. To assay induction of competence  
by synthetic competence stimulating peptide (SCSP), the cultures were incubated for 30  
min and a final concentration of 500 ng/ml of SCSP was added to each well. After a 30  
min incubation, equal amounts of DNA was added to each well (1 µg/ml plasmid or 10  
µg/ml of chromosomal DNA) and incubation continued for another 2 h. Fluid medium  
25 was then removed from individual wells and the wells were washed once with PBS  
buffer. Biofilm cells were collected into 2 ml PBS buffer by gentle sonication or by  
washing the wells using a pipette. The samples were centrifuged at 12,000 X g for 5 min.  
Both biofilm and planktonic cells were resuspended into 200 µl of fresh medium and  
were immediately spread on THYE agar plates plus appropriate antibiotics. For the  
30 biofilms developed in the chemostat, rods with biofilm cells were removed and placed  
into 2 ml of pre-warmed, fresh THYE-HS medium for 30 min incubation. Transformation  
was then initiated by using the same methods as described above. The planktonic cells  
were also removed to compare the transformation frequency. After completion of the  
transformation procedures, both biofilm and planktonic cells were spread on THYE agar

plus appropriate antibiotic. Transformation frequency was assessed after 48-h incubation. Transformation frequency was expressed as the number of transformants per  $\mu$ g DNA per viable recipient at the time of DNA added.

5     **Donor DNA**

Both plasmid and chromosomal DNA were used as donor DNA to assay genetic transformation in this study. Plasmid DNA included an integrative plasmid, pVAGTFA carrying an erythromycin resistance ( $Em^r$ ) determinant and a fragment of the *S. mutans* *gtfA* gene. The replicative plasmid, pDL289 carrying a kanamycin resistance gene ( $Km^r$ ) was also used. Chromosomal DNA harboring an  $Em^r$  gene was prepared from a recombinant *S. mutans* strain harboring a chromosomally integrated copy of pVAGTFA.

---

15     The present invention has been described in detail and with particular reference to the preferred embodiments; however, it will be understood by one having ordinary skill in the art that changes can be made without departing from the spirit and scope thereof. For example, where the application refers to peptides, it is clear that polypeptides may often be used. Likewise, where a gene is described in the application, it is clear that nucleic acid molecules or gene fragments may often be used.

20     All publications (including GenBank entries), patents and patent applications are incorporated by reference in their entirety to the same extent as if each individual publication, patent or patent application was specifically and individually indicated to be incorporated by reference in its entirety.



## SEQUENCE LISTING

&lt;110&gt; CVITKOVITCH, Dennis

5 <120> SIGNAL PEPTIDES NUCLEIC ACID MOLECULES AND METHODS FOR TREATMENT  
OF CARIES

&lt;130&gt; 311/0003

10 &lt;140&gt; 2 302 861

&lt;141&gt; 2000-04-10

&lt;160&gt; 28

15

&lt;170&gt; PatentIn version 3.0

&lt;210&gt; 1

20

&lt;211&gt; 141

&lt;212&gt; DNA

&lt;213&gt; Streptococcus mutans

25

&lt;220&gt;

&lt;221&gt; CDS

30 &lt;222&gt; (1)..(141)

&lt;400&gt; 1

atg aaa aaa aca cta tca tta aaa aat gac ttt aaa gaa att aag act 48

Met Lys Lys Thr Leu Ser Leu Lys Asn Asp Phe Lys Glu Ile Lys Thr

35

1

5

10

15

gat gaa tta gag att atc att ggc gga agc gga agc cta tca aca ttt 96

Asp Glu Leu Glu Ile Ile Ile Gly Gly Ser Gly Ser Leu Ser Thr Phe

20

25

30

40

ttc cgg ctg ttt aac aga agt ttt aca caa gct ttg gga aaa taa 141

Phe Arg Leu Phe Asn Arg Ser Phe Thr Gln Ala Leu Gly Lys

35

40

45

45

&lt;210&gt; 2

&lt;211&gt; 46

50

&lt;212&gt; PRT

&lt;213&gt; Streptococcus mutans

55

&lt;400&gt; 2

Met Lys Lys Thr Leu Ser Leu Lys Asn Asp Phe Lys Glu Ile Lys Thr  
 1 5 10 15

5 Asp Glu Leu Glu Ile Ile Ile Gly Gly Ser Gly Ser Leu Ser Thr Phe  
 20 25 30

10 Phe Arg Leu Phe Asn Arg Ser Phe Thr Gln Ala Leu Gly Lys  
 35 40 45

<210> 3

15 <211> 1326

<212> DNA

<213> Streptococcus mutans

20

<220>

25 <221> CDS

<222> (1) .. (1326)

30

<400> 3

atg aat gaa gcc tta atg ata ctt tca aat ggt tta tta act tat cta 48  
 Met Asn Glu Ala Leu Met Ile Leu Ser Asn Gly Leu Leu Thr Tyr Leu  
 1 5 10 15

35 acc gtt cta ttt ctc ttg ttt cta ttt tct aag gta agt aat gtc act 96  
 Thr Val Leu Phe Leu Leu Phe Leu Phe Ser Lys Val Ser Asn Val Thr  
 20 25 30

40 tta tcg aaa aag gaa tta act ctt ttt tcg ata agc aat ttt ctg ata 144  
 Leu Ser Lys Lys Glu Leu Thr Leu Phe Ser Ile Ser Asn Phe Leu Ile  
 35 40 45

45 atg att gct gtt acg atg gtg aac gta aac ctg ttt tat cct gca gag 192  
 Met Ile Ala Val Thr Met Val Asn Val Asn Leu Phe Tyr Pro Ala Glu  
 50 55 60

cct ctt tat ttt ata gct tta tca att tat ctt aat aga cag aat agt 240  
 Pro Leu Tyr Phe Ile Ala Leu Ser Ile Tyr Leu Asn Arg Gln Asn Ser  
 50 65 70 75 80

ctt tct cta aat ata ttt tat ggt ctg ctg cct gtt gcc agt tct gac 288  
 Leu Ser Leu Asn Ile Phe Tyr Gly Leu Leu Pro Val Ala Ser Ser Asp  
 85 90 95

55 ttg ttt agg cgg gca atc ata ttc ttt atc ttg gat gga act caa gga 336  
 Leu Phe Arg Arg Ala Ile Ile Phe Phe Ile Leu Asp Gly Thr Gln Gly

	100	105	110	
5	att gta atg ggc agt agc att ata acc acc tat atg atc gag ttt gca Ile Val Met Gly Ser Ser Ile Ile Thr Thr Tyr Met Ile Glu Phe Ala 115 120 125			384
10	gga ata gcg cta agt tac ctc ttt ctc agt gtg ttc aat gtt gat att Gly Ile Ala Leu Ser Tyr Leu Phe Leu Ser Val Phe Asn Val Asp Ile 130 135 140			432
15	ggt cga ctt aaa gat agt ttg acc aag atg aag gtc aaa aaa cgc ttg Gly Arg Leu Lys Asp Ser Leu Thr Lys Met Lys Val Lys Lys Arg Leu 145 150 155 160			480
20	att cca atg aat att act atg ctt cta tac tac ctt tta ata cag gta Ile Pro Met Asn Ile Thr Met Leu Leu Tyr Tyr Leu Leu Ile Gln Val 165 170 175			528
25	ttg tat gtt ata gag agt tat aat gtg ata ccg act tta aaa ttt cgt Leu Tyr Val Ile Glu Ser Tyr Asn Val Ile Pro Thr Leu Lys Phe Arg 180 185 190			576
30	aaa ttt gtc gtt att gtc tat ctt att tta ttt ttg att ctg atc tca Lys Phe Val Val Ile Val Tyr Leu Ile Leu Phe Leu Ile Leu Ile Ser 195 200 205			624
35	ttt tta agc caa tat acc aaa caa aag gtt caa aat gag ata atg gca Phe Leu Ser Gln Tyr Thr Lys Gln Lys Val Gln Asn Glu Ile Met Ala 210 215 220			672
40	caa aag gaa gct cag att cga aat atc acc cag tat agt cag caa ata Gln Lys Glu Ala Gln Ile Arg Asn Ile Thr Gln Tyr Ser Gln Gln Ile 225 230 235 240			720
45	gaa tct ctt tac aag gat att cga agt ttc cgc cat gat tat ctg aat Glu Ser Leu Tyr Lys Asp Ile Arg Ser Phe Arg His Asp Tyr Leu Asn 245 250 255			768
50	att tta act agc ctc aga tta ggc att gaa aat aaa gat tta gct agt Ile Leu Thr Ser Leu Arg Leu Gly Ile Glu Asn Lys Asp Leu Ala Ser 260 265 270			816
55	att gaa aag att tac cat caa atc tta gaa aaa aca gga cat caa ttg Ile Glu Lys Ile Tyr His Gln Ile Leu Glu Lys Thr Gly His Gln Leu 275 280 285			864
60	cag gat acc cgt tat aat atc ggc cat cta gct aat att caa aac gat Gln Asp Thr Arg Tyr Asn Ile Gly His Leu Ala Asn Ile Gln Asn Asp 290 295 300			912
65	gct gtc aag ggt atc ttg tca gca aaa atc tta gaa gct cag aat aaa Ala Val Lys Gly Ile Leu Ser Ala Lys Ile Leu Glu Ala Gln Asn Lys 305 310 315 320			960
70	aag att gct gtc aat gta gaa gtc tca agt aaa ata caa ctg cct gag Lys Ile Ala Val Asn Val Glu Val Ser Ser Lys Ile Gln Leu Pro Glu 325 330 335			1008

atg gag ttg ctt gat ttc att acc ata ctt tct atc ttg tgt gat aat 1056  
 Met Glu Leu Leu Asp Phe Ile Thr Ile Leu Ser Ile Leu Cys Asp Asn  
 340 345 350

5 gcc att gag gct gct ttc gaa tca tta aat cct gaa att cag tta gcc 1104  
 Ala Ile Glu Ala Ala Phe Glu Ser Leu Asn Pro Glu Ile Gln Leu Ala  
 355 360 365

10 ttt ttt aag aaa aat ggc agt ata gtc ttt atc att cag aat tcc acc 1152  
 Phe Phe Lys Lys Asn Gly Ser Ile Val Phe Ile Ile Gln Asn Ser Thr  
 370 375 380

15 aaa gaa aaa caa ata gat gtg agt aaa att ttt aaa gaa aac tat tcc 1200  
 Lys Glu Lys Gln Ile Asp Val Ser Lys Ile Phe Lys Glu Asn Tyr Ser  
 385 390 395 400

20 act aaa ggc tcc aat cgc ggt att ggt tta gca aag gtg aat cat att 1248  
 Thr Lys Gly Ser Asn Arg Gly Ile Gly Leu Ala Lys Val Asn His Ile  
 405 410 415

25 ctt gaa cat tat ccc aaa acc agt tta caa aca agc aat cat cat cat 1296  
 Leu Glu His Tyr Pro Lys Thr Ser Leu Gln Thr Ser Asn His His His  
 420 425 430

tta ttc aag caa ctc cta ata ata aaa tag 1326  
 Leu Phe Lys Gln Leu Leu Ile Ile Lys  
 435 440

30 <210> 4  
 <211> 441

35 <212> PRT  
 <213> Streptococcus mutans

40 <400> 4

Met Asn Glu Ala Leu Met Ile Leu Ser Asn Gly Leu Leu Thr Tyr Leu  
 1 5 10 15

45 Thr Val Leu Phe Leu Leu Phe Leu Phe Ser Lys Val Ser Asn Val Thr  
 20 25 30

50 Leu Ser Lys Lys Glu Leu Thr Leu Phe Ser Ile Ser Asn Phe Leu Ile  
 35 40 45

55 Met Ile Ala Val Thr Met Val Asn Val Asn Leu Phe Tyr Pro Ala Glu  
 50 55 60

5 Pro Leu Tyr Phe Ile Ala Leu Ser Ile Tyr Leu Asn Arg Gln Asn Ser  
 65 70 75 80  
 Leu Ser Leu Asn Ile Phe Tyr Gly Leu Leu Pro Val Ala Ser Ser Asp  
 85 90 95  
 10 Leu Phe Arg Arg Ala Ile Ile Phe Phe Ile Leu Asp Gly Thr Gln Gly  
 100 105 110  
 15 Ile Val Met Gly Ser Ser Ile Ile Thr Thr Tyr Met Ile Glu Phe Ala  
 115 120 125  
 20 Gly Ile Ala Leu Ser Tyr Leu Phe Leu Ser Val Phe Asn Val Asp Ile  
 130 135 140  
 25 Gly Arg Leu Lys Asp Ser Leu Thr Lys Met Lys Val Lys Lys Arg Leu  
 145 150 155 160  
 Ile Pro Met Asn Ile Thr Met Leu Leu Tyr Tyr Leu Leu Ile Gln Val  
 165 170 175  
 30 Leu Tyr Val Ile Glu Ser Tyr Asn Val Ile Pro Thr Leu Lys Phe Arg  
 180 185 190  
 35 Lys Phe Val Val Ile Val Tyr Leu Ile Leu Phe Leu Ile Leu Ile Ser  
 195 200 205  
 40 Phe Leu Ser Gln Tyr Thr Lys Gln Lys Val Gln Asn Glu Ile Met Ala  
 210 215 220  
 45 Gln Lys Glu Ala Gln Ile Arg Asn Ile Thr Gln Tyr Ser Gln Gln Ile  
 225 230 235 240  
 Glu Ser Leu Tyr Lys Asp Ile Arg Ser Phe Arg His Asp Tyr Leu Asn  
 245 250 255  
 50 Ile Leu Thr Ser Leu Arg Leu Gly Ile Glu Asn Lys Asp Leu Ala Ser  
 260 265 270  
 55 Ile Glu Lys Ile Tyr His Gln Ile Leu Glu Lys Thr Gly His Gln Leu  
 275 280 285

Gln Asp Thr Arg Tyr Asn Ile Gly His Leu Ala Asn Ile Gln Asn Asp  
 290 295 300

5 Ala Val Lys Gly Ile Leu Ser Ala Lys Ile Leu Glu Ala Gln Asn Lys  
 305 310 315 320

10 Lys Ile Ala Val Asn Val Glu Val Ser Ser Lys Ile Gln Leu Pro Glu  
 325 330 335

15 Met Glu Leu Leu Asp Phe Ile Thr Ile Leu Ser Ile Leu Cys Asp Asn  
 340 345 350

20 Ala Ile Glu Ala Ala Phe Glu Ser Leu Asn Pro Glu Ile Gln Leu Ala  
 355 360 365

Phe Phe Lys Lys Asn Gly Ser Ile Val Phe Ile Ile Gln Asn Ser Thr  
 370 375 380

25 Lys Glu Lys Gln Ile Asp Val Ser Lys Ile Phe Lys Glu Asn Tyr Ser  
 385 390 395 400

30 Thr Lys Gly Ser Asn Arg Gly Ile Gly Leu Ala Lys Val Asn His Ile  
 405 410 415

35 Leu Glu His Tyr Pro Lys Thr Ser Leu Gln Thr Ser Asn His His His  
 420 425 430

40 Leu Phe Lys Gln Leu Leu Ile Ile Lys  
 435 440

<210> 5  
 <211> 750

45 <212> DNA  
 <213> Streptococcus mutans

50 <220>  
 <221> CDS

55 <222> (1)..(750)

<400> 5		
5	atg att tct att ttt gta ttg gaa gat gat ttt tta caa caa gga cgt Met Ile Ser Ile Phe Val Leu Glu Asp Asp Phe Leu Gln Gln Gly Arg	48
	1 5 10 15	
10	ctt gaa acc acc att gca gct atc atg aaa gaa aaa aat tgg tct tat Leu Glu Thr Thr Ile Ala Ala Ile Met Lys Glu Lys Asn Trp Ser Tyr	96
	20 25 30	
15	aaa gaa ttg act att ttt gga aaa cca caa caa ctt att gac gct atc Lys Glu Leu Thr Ile Phe Gly Lys Pro Gln Gln Leu Ile Asp Ala Ile	144
	35 40 45	
20	cct gaa aag ggc aat cac cag att ttc ttt ttg gat att gaa atc aaa Pro Glu Lys Gly Asn His Gln Ile Phe Phe Leu Asp Ile Glu Ile Lys	192
	50 55 60	
25	aaa gag gaa aag aaa gga ctg gaa gta gcc aat cag att aga cag cat Lys Glu Glu Lys Lys Gly Leu Glu Val Ala Asn Gln Ile Arg Gln His	240
	65 70 75 80	
30	aat cct agt gca gtt att gtc ttt gtc acg aca cat tct gag ttt atg Asn Pro Ser Ala Val Ile Val Phe Val Thr Thr His Ser Glu Phe Met	288
	85 90 95	
35	ccc ctc act ttt cag tat cag gta tct gct ttg gat ttt att gat aaa Pro Leu Thr Phe Gln Tyr Gln Val Ser Ala Leu Asp Phe Ile Asp Lys	336
	100 105 110	
40	tct ttg aat cct gag gag ttc tcc cac cgc att gaa tca gcg ctg tat Ser Leu Asn Pro Glu Glu Phe Ser His Arg Ile Glu Ser Ala Leu Tyr	384
	115 120 125	
45	tat gct atg gaa aac agc cag aag aat ggt caa tca gag gaa ctt ttt Tyr Ala Met Glu Asn Ser Gln Lys Asn Gly Gln Ser Glu Glu Leu Phe	432
	130 135 140	
50	att ttc cat tca tct gaa act cag ttt cag gtc cct ttt gct gag att Ile Phe His Ser Ser Glu Thr Gln Phe Gln Val Pro Phe Ala Glu Ile	480
	145 150 155 160	
55	ctg tat ttt gaa aca tct tca aca gcc cat aag ctc tgc ctt tat act Leu Tyr Phe Glu Thr Ser Ser Thr Ala His Lys Leu Cys Leu Tyr Thr	528
	165 170 175	
60	tat gat gaa cgg att gaa ttc tac ggc agt atg act gac att gtt aaa Tyr Asp Glu Arg Ile Glu Phe Tyr Gly Ser Met Thr Asp Ile Val Lys	576
	180 185 190	
65	atg gat aag aga ctt ttt cag tgc cat cgc tct ttt att gtc aat cct Met Asp Lys Arg Leu Phe Gln Cys His Arg Ser Phe Ile Val Asn Pro	624
	195 200 205	
70	gcc aat att acc cgt att gat cgg aaa aaa cgc ttg gcc tat ttt cga Ala Asn Ile Thr Arg Ile Asp Arg Lys Lys Arg Leu Ala Tyr Phe Arg	672

```

      210      215      220
aat aat aag tct tgt ctt att tca cga act aag tta aca aaa ctg aga      720
Asn Asn Lys Ser Cys Leu Ile Ser Arg Thr Lys Leu Thr Lys Leu Arg
5 225      230      235      240

gct gtg att gct gat caa agg aga gca aaa      750
Ala Val Ile Ala Asp Gln Arg Arg Ala Lys
      245      250
-10-
<210> 6
<211> 250
15 <212> PRT
<213> Streptococcus mutans
20
<400> 6
25 Met Ile Ser Ile Phe Val Leu Glu Asp Asp Phe Leu Gln Gln Gly Arg
1      5      10      15

Leu Glu Thr Thr Ile Ala Ala Ile Met Lys Glu Lys Asn Trp Ser Tyr
      20      25      30
30

Lys Glu Leu Thr Ile Phe Gly Lys Pro Gln Gln Leu Ile Asp Ala Ile
      35      40      45
35

Pro Glu Lys Gly Asn His Gln Ile Phe Phe Leu Asp Ile Glu Ile Lys
      50      55      60
40

Lys Glu Glu Lys Lys Gly Leu Glu Val Ala Asn Gln Ile Arg Gln His
65      70      75      80
45

Asn Pro Ser Ala Val Ile Val Phe Val Thr Thr His Ser Glu Phe Met
      85      90      95
50

Pro Leu Thr Phe Gln Tyr Gln Val Ser Ala Leu Asp Phe Ile Asp Lys
      100      105      110
55

Ser Leu Asn Pro Glu Glu Phe Ser His Arg Ile Glu Ser Ala Leu Tyr
      115      120      125
60

Tyr Ala Met Glu Asn Ser Gln Lys Asn Gly Gln Ser Glu Glu Leu Phe
      130      135      140

```



5 Ile Phe His Ser Ser Glu Thr Gln Ph Gln Val Pro Phe Ala Glu Ile  
 145 150 155 160  
 Leu Tyr Phe Glu Thr Ser Ser Thr Ala His Lys Leu Cys Leu Tyr Thr  
 165 170 175  
 10 Tyr Asp Glu Arg Ile Glu Phe Tyr Gly Ser Met Thr Asp Ile Val Lys  
 180 185 190  
 15 Met Asp Lys Arg Leu Phe Gln Cys His Arg Ser Phe Ile Val Asn Pro  
 195 200 205  
 20 Ala Asn Ile Thr Arg Ile Asp Arg Lys Lys Arg Leu Ala Tyr Phe Arg  
 210 215 220  
 25 Asn Asn Lys Ser Cys Leu Ile Ser Arg Thr Lys Leu Thr Lys Leu Arg  
 225 230 235 240  
 Ala Val Ile Ala Asp Gln Arg Arg Ala Lys  
 245 250  
 30 <210> 7  
 <211> 46  
 35 <212> PRT  
 <213> Streptococcus mutans  
 40 <220>  
 <221> PEPTIDE  
 45 <222> (1)..(46)  
 50 <400> 7  
 Met Lys Lys Thr Pro Ser Leu Lys Asn Asp Phe Lys Glu Ile Lys Thr  
 1 5 10 15  
 55 Asp Glu Leu Glu Ile Ile Ile Gly Gly Ser Gly Ser Leu Ser Thr Phe  
 20 25 30  
 Phe Arg Leu Phe Asn Arg Ser Phe Thr Gln Ala Leu Gly Lys

35 40 45

<210> 8

5 <211> 46

<212> PRT

10 <213> Streptococcus mutans

<220>

15 <221> PEPTIDE

<222> (1)..(46)

20 <400> 8

Met Lys Lys Thr Leu Ser Leu Lys Asn Asp Phe Lys Glu Ile Lys Thr  
 1 5 10 15

25 Asp Glu Leu Glu Ile Ile Ile Gly Gly Ser Gly Ser Leu Ser Thr Phe  
 20 25 30

30 Phe Arg Leu Phe Asn Arg Ser Phe Thr Gln Ala Leu Gly Lys  
 35 40 45

<210> 9

<211> 46

35 <212> PRT

<213> Streptococcus mutans

40 <220>

<221> PEPTIDE

45 <222> (1)..(46)

50 <400> 9

Met Lys Lys Thr Leu Ser Leu Lys Asn Asp Phe Lys Glu Ile Lys Thr  
 1 5 10 15

55 Asp Glu Leu Glu Ile Ile Ile Gly Gly Ser Gly Ser Leu Ser Thr Phe  
 20 25 30

Phe Arg Leu Phe Asn Arg Ser Phe Thr Gln Ala Leu Gly Lys  
           35                  40                  45

5   <210> 10  
      <211> 43  
      <212> PRT

---

10   <213> Streptococcus mutans

15   <220>  
      <221> PEPTIDE  
      <222> (1)..(43)

20   <400> 10

25   Met Lys Lys Thr Leu Ser Leu Lys Asn Asp Phe Lys Glu Ile Lys Thr  
      1                  5                  10                  15  
      Asp Glu Leu Glu Ile Ile Ile Gly Gly Ser Gly Thr Leu Ser Thr Phe  
                   20                  25                  30

30   Phe Arg Leu Phe Asn Arg Ser Phe Thr Gln Ala  
           35                  40

35   <210> 11  
      <211> 46  
      <212> PRT

40   <213> Streptococcus mutans

45   <220>  
      <221> PEPTIDE  
      <222> (1)..(46)

50   <400> 11

55   Met Lys Lys Thr Leu Ser Leu Lys Asn Asp Phe Lys Glu Ile Lys Thr  
      1                  5                  10                  15  
      Asp Glu Leu Glu Ile Ile Ile Gly Gly Ser Gly Ser Leu Ser Thr Phe  
                   20                  25                  30

Phe Arg Leu Phe Asn Arg Ser Phe Thr Gln Ala Leu Gly Lys  
           35                          40                          45

5 <210> 12

<211> 46

<212> PRT

10

<213> Streptococcus mutans

15 <220>

<221> PEPTIDE

<222> (1)..(46)

20

<400> 12

25 Met Lys Lys Thr Leu Ser Leu Lys Asn Asp Phe Lys Glu Ile Lys Thr  
     1                          5                          10                          15

Asp Glu Leu Glu Ile Ile Ile Gly Gly Ser Gly Ser Leu Ser Thr Phe  
                           20                          25                          30

30

Phe Arg Leu Phe Asn Arg Ser Phe Thr Gln Ala Leu Gly Lys  
           35                          40                          45

<210> 13

35

<211> 46

<212> PRT

40

<213> Streptococcus mutans

<220>

45

<221> PEPTIDE

<222> (1)..(46)

50

<400> 13

55 Met Lys Lys Thr Leu Ser Leu Lys Asn Asp Phe Lys Glu Ile Lys Thr  
     1                          5                          10                          15

Asp Glu Leu Glu Ile Ile Ile Gly Gly Ser Gly Ser Leu Ser Thr Phe

	20	25	30
	Phe Arg Leu Phe Asn Arg Ser	Phe Thr Gln Ala Leu Gly Lys	
	35	40	45
5	<210> 14		
	<211> 21		
10	<212> PRT		
	<213> synthetic construct		
15	<220>		
	<221> PEPTIDE		
20	<222> (1) .. (21)		
	<400> 14		
25	Ser Gly Ser Leu Ser Thr Phe Phe Arg Leu Phe Asn Arg Ser Phe Thr		
	1 5 10 15		
30	Gln Ala Leu Gly Lys		
	20		
	<210> 15		
	<211> 19		
35	<212> DNA		
	<213> synthetic construct		
40	<220>		
	<221> PRIMER		
45	<222> (1) .. (19)		
50	<400> 15		
	agttttttgt ctggtgcg		19
55	<210> 16		
	<211> 20		

<212> DNA  
<213> synthetic construct  
5  
<220>  
<221> PRIMER  
10  
<222> (1) .. (20)  
  
15 <400> 16  
tccactaaag gctccaatcg  
20  
<210> 17  
<211> 24  
<212> DNA  
25 <213> synthetic construct  
  
30 <220>  
<221> PRIMER  
<222> (1) .. (24)  
35  
<400> 17  
cgctaagtta cctctttctc agtg  
40  
<210> 18  
<211> 21  
45 <212> DNA  
<213> synthetic construct  
  
50 <220>  
<221> PRIMER  
55 <222> (1) .. (21)

20

24

5      <400> 18  
      gcttcctttt gtgccattat c  
21

      <210> 19

10     <211> 21

      <212> DNA

      <213> synthetic construct

15

      <220>

20     <221> PRIMER

      <222> (1) .. (21)

25     <400> 19  
      cctgaaaagg gcaatcacca g  
21

30     <210> 20

      <211> 22

      <212> DNA

35     <213> synthetic construct

40     <220>

      <221> PRIMER

      <222> (1) .. (22)

45

50     <400> 20  
      gcgatggcac tgaaaaagtc tc  
22

55     <210> 21

      <211> 2557

      <212> DNA

<213> Streptococcus mutans

5 <220>

<221> misc\_feature

10 <222> (1)..(2557)

<400> 21  
 15 acattatgtg tcctaaggaa aatattactt tttcaagaaa atccatgatt ttttcataaa 60  
 aaatagtata ctaattataa tcaaaaaaag gagatataaa atgaaaaaaa cactatcatt 120  
 aaaaaatgac tttaaagaaa ttaagactga tgaattagag attatcattg gcggaagcgg 180  
 20 aagcctatca acatttttcc ggctgtttta cagaagtttt acacaagctt tgggaaaata 240  
 agataggcta acattggaat aaaacaaggc tggatttatt attccagcct ttttaaattgt 300  
 25 aaaataaaaa tacagggtta aataatcaag tgtgctgtcg tggatgagaa gataaaacta 360  
 tctcttagag aataggcctc ctctatttta ttattaggag ttgcttgaat aaatgatgat 420  
 gattgcttgt ttgtaaactg gttttgggat aatgttcaag aatatgattc acctttgcta 480  
 30 aaccaatacc gcgattggag ccttttagtgg aatagttttc tttaaaaatt ttactcacat 540  
 ctattttgtt ttctttgggtg gaattctgaa tgataaagac tatactgcca tttttcttaa 600  
 aaaaggctaa ctgaatttca ggatttaatg attcgaaagc agcctcaatg gcattatcac 660  
 35 acaagataga aagtatggta atgaaatcaa gcaactccat ctcaggcagt tgtattttac 720  
 ttgagacttc tacattgaca gcaatctttt tattctgagc ttctaagatt tttgctgaca 780  
 40 agataccctt gacagcatcg ttttgaatat tagctagatg gccgatatta taacgggtat 840  
 cctgcaattg atgtcctggt ttttctaaga tttgatggta aatcttttca atactagcta 900  
 aatctttatt ttcaatgcct aatctgaggc tagttaaaat attcagataa tcatggcgga 960  
 45 aacttcgaat atccttgtaa agagattcta tttgctgact atactgggtg atatttcgaa 1020  
 tctgagcttc cttttgtgcc attatctcat tttgaacctt ttgtttggta tattggctta 1080  
 50 aaaatgagat cagaatcaaa aataaaataa gatagacaat aacgacaaat ttacgaaatt 1140  
 ttaaagtcgg tatcacatta taactctcta taacatacaa tacctgtatt aaaaggtagt 1200  
 atagaagcat agtaatatc attggaatca agcgtttttt gaccttcac ttggtcaaac 1260  
 55 tatctttaag tcgaccaata tcaacattga acacactgag aaagaggtaa cttagcgcta 1320



ttcttgcaaa ctcgatcata taggtgggta taatgctact gccattaca attccttgag 1380  
 ttccatccaa gataaagaat atgattgccc gcctaaacaa gtcagaactg gcaacaggca 1440  
 5 gcagaccata aaatatatatt agagaaagac tattctgtct attaagataa attgataaag 1500  
 ctataaaata aagaggctct gcaggataaa acagggtttac gttcaccatc gtaacagcaa 1560  
 10 tcattatcag aaaattgctt atcgaaaaaa gagttaattc ctttttcgat aaagtgcacat 1620  
 tacttacctt agaaaataga aacaagagaa atagaacggg tagataagtt aataaaccat 1680  
 ttgaaagtat cattaaggct tcattcattt tgctctcctt tgatcagcaa tcacagctct 1740  
 15 cagttttggt aacttagttc gtgaaataag acaagactta ttatttcgaa aataggccaa 1800  
 gcgttttttc cgatcaatac gggtaatatt ggcaggattg acaataaaaag agcgatggca 1860  
 20 ctgaaaaagt ctcttatcca ttttaacaat gtcagtcata ctgccgtaga attcaatccg 1920  
 ttcatacataa gtataaaggc agagcttatg ggctgttgaa gatgtttcaa aatacagaat 1980  
 ctcagcaaaa gggacctgaa actgagtttc agatgaatgg aaaataaaaaa gttcctctga 2040  
 25 ttgaccattc ttctggctgt tttccatagc ataatacagc gctgattcaa tgcgggtggga 2100  
 gaactcctca ggattcaaag atttatcaat aaaatccaaa gcagatacct gatactgaaa 2160  
 30 agtgaggggc ataaactcag aatgtgtcgt gacaaagaca ataactgcac taggattatg 2220  
 ctgtctaate tgattggcta cttccagtc tttcttttcc tcttttttga tttcaatate 2280  
 caaaaagaaa atctggatgat tgcccttttc agggatagcg tcaataagtt gttgtgggtt 2340  
 35 tccaaaaata gtcaattctt tataagacca atttttttct ttcattgatag ctgcaatggt 2400  
 gggttcaaga cgtccttggt gtaaaaaatc atcttccaat acaaaaatag aaatcattat 2460  
 40 ttctccttta atcttctatt taggttagct gattaacact atacacagaa aaggtataaa 2520  
 acgatatcac tcaataaaat ctactaactt aataacc 2557  
  
 45 <210> 22  
 <211> 480  
 <212> DNA  
 50 <213> Streptococcus mutans  
  
 55 <220>  
 <221> CDS

&lt;222&gt; (1)..(408)

5 <400> 22  
 atg gaa gaa gat ttt gaa att gtt ttt aat aag gtt aag cca att gta 48  
 Met Glu Glu Asp Phe Glu Ile Val Phe Asn Lys Val Lys Pro Ile Val  
 1 5 10 15

10 tgg aaa tta agc cgt tat tac ttt att aaa atg tgg act cgt gaa gat 96  
 Trp Lys Leu Ser Arg Tyr Tyr Phe Ile Lys Met Trp Thr Arg Glu Asp  
 20 25 30

15 tgg caa caa gag gga atg ttg att ttg cac caa tta tta agg gaa cat 144  
 Trp Gln Gln Glu Gly Met Leu Ile Leu His Gln Leu Leu Arg Glu His  
 35 40 45

20 cca gaa tta gaa gag gat gat aca aaa ttg tat atc tat ttt aag aca 192  
 Pro Glu Leu Glu Glu Asp Asp Thr Lys Leu Tyr Ile Tyr Phe Lys Thr  
 50 55 60

25 cgt ttt tct aat tac att aaa gat gtt ttg cgt cag caa gaa agt cag 240  
 Arg Phe Ser Asn Tyr Ile Lys Asp Val Leu Arg Gln Gln Glu Ser Gln  
 65 70 75 80

aaa cgt cgt ttt aat aga atg tct tat gaa gaa gtc ggt gag att gaa 288  
 Lys Arg Arg Phe Asn Arg Met Ser Tyr Glu Glu Val Gly Glu Ile Glu  
 85 90 95

30 cac tgt ttg tca agt ggc ggt atg caa ttg gat gaa tat att tta ttt 336  
 His Cys Leu Ser Ser Gly Gly Met Gln Leu Asp Glu Tyr Ile Leu Phe  
 100 105 110

35 cgt gat agt ttg ctt gca tat aaa caa ggt ctg agt act gaa aag caa 384  
 Arg Asp Ser Leu Leu Ala Tyr Lys Gln Gly Leu Ser Thr Glu Lys Gln  
 115 120 125

40 gag ctg ttt gag cgc ttg gta gca ggagagcact ttttggaag gcaaagtatg 438  
 Glu Leu Phe Glu Arg Leu Val Ala  
 130 135

ctgaaagatt tacgtaaaaa attaagtgat ttttaaggaaa aa 480

45 <210> 23  
 <211> 136  
 <212> PRT  
 50 <213> Streptococcus mutans

55 <400> 23  
 Met Glu Glu Asp Phe Glu Ile Val Phe Asn Lys Val Lys Pro Ile Val

1 5 10 15

5 Trp Lys Leu Ser Arg Tyr Tyr Phe Ile Lys Met Trp Thr Arg Glu Asp  
20 25 30

10 Trp Gln Gln Glu Gly Met Leu Ile Leu His Gln Leu Leu Arg Glu His  
35 40 45

15 Pro Glu Leu Glu Glu Asp Asp Thr Lys Leu Tyr Ile Tyr Phe Lys Thr  
50 55 60

20 Arg Phe Ser Asn Tyr Ile Lys Asp Val Leu Arg Gln Gln Glu Ser Gln  
65 70 75 80

25 Lys Arg Arg Phe Asn Arg Met Ser Tyr Glu Glu Val Gly Glu Ile Glu  
85 90 95

30 His Cys Leu Ser Ser Gly Gly Met Gln Leu Asp Glu Tyr Ile Leu Phe  
100 105 110

35 Arg Asp Ser Leu Leu Ala Tyr Lys Gln Gly Leu Ser Thr Glu Lys Gln  
115 120 125

40 Glu Leu Phe Glu Arg Leu Val Ala  
130 135

45 <210> 24  
<211> 680

50 <212> DNA  
<213> Streptococcus mutans

55 <220>  
<221> misc\_feature  
<222> (1)..(680)

60 <400> 24  
gtaaataaaa cagccagtta agatgggaca tttatgtcct gttcttaaag tctttttcgt 60  
tttataataa ttttattata aaaggaggtc atcgtaatag atggaagaag attttgaat 120

	tgtttttaat aaggttaagc caattgtatg gaaattaagc cgttattact ttattaaaaat	180
5	gtggactcgt gaagattggc aacaagaggg aatgttgatt ttgcaccaat tattaaggga	240
	acatccagaa ttagaagagg atgatacaaa attgtatatc tattttaaga cacgtttttc	300
	taattacatt aaagatgttt tgcgtcagca agaaagtcag aaacgtcgtt ttaatagaat	360
10	gtcttatgaa gaagtcggtg agattgaaca ctgtttgtca agtggcggtg tgcaattgga	420
	tgaatatatt ttatttcgtg atagtttgct tgcataataa caaggctctga gtactgaaaa	480
15	gcaagagctg tttgagcgct tggtagcagg agagcacttt ttgggaaggc aaagtatgct	540
	gaaagattta cgtaaaaaat taagtgtatt taaggaaaaa tagttaaaaa gggaaagaat	600
	ggaacatgtg attgtaccat tctttttggg tgaaaattaa gaaaagttat tataaattat	660
20	tggtttaaca tgccatatta	680

<210> 25

25 <211> 2280

**<212> DNA**

30 <213> Streptococcus mutans

**<220>**

35 <221> CDS

<222> (1) . . (2280)

40

atg aaa caa gtt att tat gtt gtt tta atc gtc ata gcc gtt aac att 48  
Met Lys Gln Val Ile Tyr Val Val Leu Ile Val Ile Ala Val Asn Ile  
1 5 10 15

ctc tta gag att atc aaa aga gta aca aaa agg gga ggg aca gtt tcg  
Leu Leu Glu Ile Ile Lys Arg Val Thr Lys Arg Gly Gly Thr Val Ser

20                  25                  30

50 tca tct aat cct tta cca gat ggg cag tct aag ttg ttt tgg cgc aga 144  
Ser Ser Asn Pro Leu Pro Asp Gly Gln Ser Lys Leu Phe Trp Arg Arg  
35 40 45

55 cat tat aag cta gta cct cag att gat acc aga gac tgt ggg ccg gca 192  
His Tyr Lys Leu Val Pro Gln Ile Asp Thr Arg Asp Cys Gly Pro Ala  
50 55 60

	gtg ctg gca tct gtt gca aag cat tac gga tct aat tac tct atc gct	240
	Val Leu Ala Ser Val Ala Lys His Tyr Gly Ser Asn Tyr Ser Ile Ala	
	65 70 75 80	
5	tat ctg cgg gaa ctc tca aag act aac aag cag gga aca aca gct ctt	288
	Tyr Leu Arg Glu Leu Ser Lys Thr Asn Lys Gln Gly Thr Thr Ala Leu	
	85 90 95	
10	ggc att gtt gaa gct gct aaa aag tta ggc ttt gaa aca cgc tct atc	336
	Gly Ile Val Glu Ala Ala Lys Lys Leu Gly Phe Glu Thr Arg Ser Ile	
	100 105 110	
15	aag gcg gat atg acg ctt ttt gat tat aat gat ttg acc tat cct ttt	384
	Lys Ala Asp Met Thr Leu Phe Asp Tyr Asn Asp Leu Thr Tyr Pro Phe	
	115 120 125	
20	atc gtc cat gtg att aaa gga aaa cgt ctg cag cat tat tat gtc gtc	432
	Ile Val His Val Ile Lys Gly Lys Arg Leu Gln His Tyr Tyr Val Val	
	130 135 140	
	tat ggc agc cag aat aat cag ctg att att gga gat cct gat cct tca	480
	Tyr Gly Ser Gln Asn Asn Gln Leu Ile Ile Gly Asp Pro Asp Pro Ser	
	145 150 155 160	
25	gtt aag gtg act agg atg agt aag gaa cgc ttt caa tca gag tgg aca	528
	Val Lys Val Thr Arg Met Ser Lys Glu Arg Phe Gln Ser Glu Trp Thr	
	165 170 175	
30	ggc ctt gca att ttc cta gct cct cag cct aac tat aag cct cat aaa	576
	Gly Leu Ala Ile Phe Leu Ala Pro Gln Pro Asn Tyr Lys Pro His Lys	
	180 185 190	
35	ggc gaa aaa aat ggt ttg tct aat ttc ttc ccg ttg atc ttt aag cag	624
	Gly Glu Lys Asn Gly Leu Ser Asn Phe Phe Pro Leu Ile Phe Lys Gln	
	195 200 205	
40	aaa gct ttg atg act tat att atc ata gct agc ttg att gtg acg ctc	672
	Lys Ala Leu Met Thr Tyr Ile Ile Ile Ala Ser Leu Ile Val Thr Leu	
	210 215 220	
	att gat att gtc gga tca tac tat ctc caa gga ata ttg gac gag tac	720
	Ile Asp Ile Val Gly Ser Tyr Tyr Leu Gln Gly Ile Leu Asp Glu Tyr	
	225 230 235 240	
45	att cct gat cag ctg att tca act tta gga atg att acg att ggt ctg	768
	Ile Pro Asp Gln Leu Ile Ser Thr Leu Gly Met Ile Thr Ile Gly Leu	
	245 250 255	
50	ata ata acc tat att atc cag cag gtc atg gct ttt gca aaa gaa tac	816
	Ile Ile Thr Tyr Ile Ile Gln Gln Val Met Ala Phe Ala Lys Glu Tyr	
	260 265 270	
55	ctc ttg gcc gta ctc agt ttg cgt tta gtc att gat gtt atc ctg tct	864
	Leu Leu Ala Val Leu Ser Leu Arg Leu Val Ile Asp Val Ile Leu Ser	
	275 280 285	
	tat atc aaa cat att ttt acg ctt cct atg tct ttc ttt gcg aca agg	912

	Tyr Ile Lys His Ile Phe Thr Leu Pro Met Ser Phe Phe Ala Thr Arg	
	290 295 300	
5	cga aca gga gaa atc acg tct cgt ttt aca gat gcc aat cag att att Arg Thr Gly Glu Ile Thr Ser Arg Phe Thr Asp Ala Asn Gln Ile Ile	960
	305 310 315 320	
10	gat gct gta gcg tca acc atc ttt tca atc ttt tta gat atg act atg Asp Ala Val Ala Ser Thr Ile Phe Ser Ile Phe Leu Asp Met Thr Met	1008
	325 330 335	
15	gta att ttg gtt ggt ggg gtt ttg ttg gcg caa aac aat aac ctt ttc Val Ile Leu Val Gly Gly Val Leu Leu Ala Gln Asn Asn Asn Leu Phe	1056
	340 345 350	
20	ttt cta acc ttg ctc tcc att ccg att tat gcc atc att att ttt gct Phe Leu Thr Leu Leu Ser Ile Pro Ile Tyr Ala Ile Ile Ile Phe Ala	1104
	355 360 365	
25	ttc ttg aaa ccc ttt gag aaa atg aat cac gaa gtg atg gaa agc aat Phe Leu Lys Pro Phe Glu Lys Met Asn His Glu Val Met Glu Ser Asn	1152
	370 375 380	
30	gct gtg gta agt tct tct atc att gaa gat atc aat ggg atg gaa acc Ala Val Val Ser Ser Ser Ile Ile Glu Asp Ile Asn Gly Met Glu Thr	1200
	385 390 395 400	
35	att aaa tca ctc aca agt gag tcc gct cgt tat caa aac att gat agt Ile Lys Ser Leu Thr Ser Glu Ser Ala Arg Tyr Gln Asn Ile Asp Ser	1248
	405 410 415	
40	gaa ttt gtt gat tat ttg gag aaa aac ttt aag cta cac aag tat agt Glu Phe Val Asp Tyr Leu Glu Lys Asn Phe Lys Leu His Lys Tyr Ser	1296
	420 425 430	
45	gcc att caa acc gca tta aaa agc ggt gct aag ctt atc ctc aat gtt Ala Ile Gln Thr Ala Leu Lys Ser Gly Ala Lys Leu Ile Leu Asn Val	1344
	435 440 445	
50	gtc att ctc tgg tat ggc tct cgt cta gtt atg gat aat aaa atc tca Val Ile Leu Trp Tyr Gly Ser Arg Leu Val Met Asp Asn Lys Ile Ser	1392
	450 455 460	
55	gtt ggt cag ctt atc acc ttt aat gct ttg ctg tct tat ttc tca aat Val Gly Gln Leu Ile Thr Phe Asn Ala Leu Leu Ser Tyr Phe Ser Asn	1440
	465 470 475 480	
60	cca att gaa aat att atc aat ctg caa tcc aaa ctg cag tca gct cgc Pro Ile Glu Asn Ile Ile Asn Leu Gln Ser Lys Leu Gln Ser Ala Arg	1488
	485 490 495	
65	gtt gcc aat aca cgt ctt aat gag gtc tat ctt gtc gaa tct gaa ttt Val Ala Asn Thr Arg Leu Asn Glu Val Tyr Leu Val Glu Ser Glu Phe	1536
	500 505 510	
70	gaa aaa gac ggc gat tta tca gaa aat agc ttt tta gat ggt gat att Glu Lys Asp Gly Asp Leu Ser Glu Asn Ser Phe Leu Asp Gly Asp Ile	1584

	515	520	525	
5	tcg ttt gaa aat ctt tct tat aaa tat gga ttt ggg cga gat acc tta Ser Phe Glu Asn Leu Ser Tyr Lys Tyr Gly Phe Gly Arg Asp Thr Leu 530 535 540			1632
10	tca gat att aat tta tca atc aaa aaa ggc tcc aag gtc agt cta gtt Ser Asp Ile Asn Leu Ser Ile Lys Lys Gly Ser Lys Val Ser Leu Val 545 550 555 560			1680
15	gga gcc agt ggt tct ggt aaa aca act ttg gct aaa ctg att gtc aat Gly Ala Ser Gly Ser Gly Lys Thr Thr Leu Ala Lys Leu Ile Val Asn 565 570 575			1728
20	ttc tac gag cct aac aag ggg att gtt cga atc aat ggc aat gat tta Phe Tyr Glu Pro Asn Lys Gly Ile Val Arg Ile Asn Gly Asn Asp Leu 580 585 590			1776
25	aaa gtt att gat aag aca gct ttg cgg cgg cat att agc tat ttg ccg Lys Val Ile Asp Lys Thr Ala Leu Arg Arg His Ile Ser Tyr Leu Pro 595 600 605			1824
30	caa cag gcc tat gtt ttt agt ggc tct att atg gat aat ctc gtt tta Gln Gln Ala Tyr Val Phe Ser Gly Ser Ile Met Asp Asn Leu Val Leu 610 615 620			1872
35	gga gct aaa gaa gga acg agt cag gaa gac att att cgt gct tgt gaa Gly Ala Lys Glu Gly Thr Ser Gln Glu Asp Ile Ile Arg Ala Cys Glu 625 630 635 640			1920
40	att gct gaa atc cgc tgc gac att gaa caa atg cct cag ggc tat cag Ile Ala Glu Ile Arg Ser Asp Ile Glu Gln Met Pro Gln Gly Tyr Gln 645 650 655			1968
45	aca gag tta tca gat ggt gcc ggt att tct ggc ggt caa aaa cag ccg Thr Glu Leu Ser Asp Gly Ala Gly Ile Ser Gly Gly Gln Lys Gln Arg 660 665 670			2016
50	att gct tta gct agg gcc tta tta aca cag gca ccg gtt ttg att ctg Ile Ala Leu Ala Arg Ala Leu Leu Thr Gln Ala Pro Val Leu Ile Leu 675 680 685			2064
55	gat gaa gcc acc agc agt ctt gat att ttg aca gaa aag aaa att atc Asp Glu Ala Thr Ser Ser Leu Asp Ile Leu Thr Glu Lys Lys Ile Ile 690 695 700			2112
60	agc aat ctc tta cag atg acg gag aaa aca ata att ttt gtt gcc cac Ser Asn Leu Leu Gln Met Thr Glu Lys Thr Ile Ile Phe Val Ala His 705 710 715 720			2160
65	cgc tta agc att tca cag cgt act gac gaa gtc att gtc atg gat cag Arg Leu Ser Ile Ser Gln Arg Thr Asp Glu Val Ile Val Met Asp Gln 725 730 735			2208
70	gga aaa att gtt gaa caa ggc act cat aag gaa ctt tta gct aag caa Gly Lys Ile Val Glu Gln Gly Thr His Lys Glu Leu Leu Ala Lys Gln 740 745 750			2256

ggt ttc tat tat aac ctg ttt aat  
 Gly Phe Tyr Tyr Asn Leu Phe Asn  
       755                      760

2280

5

&lt;210&gt; 26

10

&lt;211&gt; 760

&lt;212&gt; PRT

&lt;213&gt; Streptococcus mutans

15

&lt;400&gt; 26

20

Met Lys Gln Val Ile Tyr Val Val Leu Ile Val Ile Ala Val Asn Ile  
   1                      5                      10                      15

25

Leu Leu Glu Ile Ile Lys Arg Val Thr Lys Arg Gly Gly Thr Val Ser  
                       20                      25                      30

30

Ser Ser Asn Pro Leu Pro Asp Gly Gln Ser Lys Leu Phe Trp Arg Arg  
                       35                      40                      45

His Tyr Lys Leu Val Pro Gln Ile Asp Thr Arg Asp Cys Gly Pro Ala  
   50                      55                      60

35

Val Leu Ala Ser Val Ala Lys His Tyr Gly Ser Asn Tyr Ser Ile Ala  
   65                      70                      75                      80

40

Tyr Leu Arg Glu Leu Ser Lys Thr Asn Lys Gln Gly Thr Thr Ala Leu  
                       85                      90                      95

45

Gly Ile Val Glu Ala Ala Lys Lys Leu Gly Phe Glu Thr Arg Ser Ile  
                       100                      105                      110

Lys Ala Asp Met Thr Leu Phe Asp Tyr Asn Asp Leu Thr Tyr Pro Phe  
                       115                      120                      125

50

Ile Val His Val Ile Lys Gly Lys Arg Leu Gln His Tyr Tyr Val Val  
   130                      135                      140

55

Tyr Gly Ser Gln Asn Asn Gln Leu Ile Ile Gly Asp Pro Asp Pro Ser  
   145                      150                      155                      160



Val Lys Val Thr Arg Met Ser Lys Glu Arg Phe Gln Ser Glu Trp Thr  
 165 170 175  
 5 Gly Leu Ala Ile Phe Leu Ala Pro Gln Pro Asn Tyr Lys Pro His Lys  
 180 185 190  
 10 Gly Glu Lys Asn Gly Leu Ser Asn Phe Phe Pro Leu Ile Phe Lys Gln  
 195 200 205  
 15 Lys Ala Leu Met Thr Tyr Ile Ile Ile Ala Ser Leu Ile Val Thr Leu  
 210 215 220  
 20 Ile Asp Ile Val Gly Ser Tyr Tyr Leu Gln Gly Ile Leu Asp Glu Tyr  
 225 230 235 240  
 25 Ile Pro Asp Gln Leu Ile Ser Thr Leu Gly Met Ile Thr Ile Gly Leu  
 245 250 255  
 30 Leu Leu Ala Val Leu Ser Leu Arg Leu Val Ile Asp Val Ile Leu Ser  
 275 280 285  
 35 Tyr Ile Lys His Ile Phe Thr Leu Pro Met Ser Phe Phe Ala Thr Arg  
 290 295 300  
 40 Arg Thr Gly Glu Ile Thr Ser Arg Phe Thr Asp Ala Asn Gln Ile Ile  
 305 310 315 320  
 45 Asp Ala Val Ala Ser Thr Ile Phe Ser Ile Phe Leu Asp Met Thr Met  
 325 330 335  
 50 Val Ile Leu Val Gly Gly Val Leu Leu Ala Gln Asn Asn Asn Leu Phe  
 340 345 350  
 55 Phe Leu Thr Leu Leu Ser Ile Pro Ile Tyr Ala Ile Ile Ile Phe Ala  
 355 360 365  
 Phe Leu Lys Pro Phe Glu Lys Met Asn His Glu Val Met Glu Ser Asn  
 370 375 380

Ala Val Val Ser Ser Ser Ile Ile Glu Asp Ile Asn Gly Met Glu Thr  
 385 390 395 400

5 Ile Lys Ser Leu Thr Ser Glu Ser Ala Arg Tyr Gln Asn Ile Asp Ser  
 405 410 415

10 Glu Phe Val Asp Tyr Leu Glu Lys Asn Phe Lys Leu His Lys Tyr Ser  
 420 425 430

15 Ala Ile Gln Thr Ala Leu Lys Ser Gly Ala Lys Leu Ile Leu Asn Val  
 435 440 445

Val Ile Leu Trp Tyr Gly Ser Arg Leu Val Met Asp Asn Lys Ile Ser  
 450 455 460

20 Val Gly Gln Leu Ile Thr Phe Asn Ala Leu Leu Ser Tyr Phe Ser Asn  
 465 470 475 480

25 Pro Ile Glu Asn Ile Ile Asn Leu Gln Ser Lys Leu Gln Ser Ala Arg  
 485 490 495

30 Val Ala Asn Thr Arg Leu Asn Glu Val Tyr Leu Val Glu Ser Glu Phe  
 500 505 510

35 Glu Lys Asp Gly Asp Leu Ser Glu Asn Ser Phe Leu Asp Gly Asp Ile  
 515 520 525

Ser Phe Glu Asn Leu Ser Tyr Lys Tyr Gly Phe Gly Arg Asp Thr Leu  
 530 535 540

40 Ser Asp Ile Asn Leu Ser Ile Lys Lys Gly Ser Lys Val Ser Leu Val  
 545 550 555 560

45 Gly Ala Ser Gly Ser Gly Lys Thr Thr Leu Ala Lys Leu Ile Val Asn  
 565 570 575

50 Phe Tyr Glu Pro Asn Lys Gly Ile Val Arg Ile Asn Gly Asn Asp Leu  
 580 585 590

Lys Val Ile Asp Lys Thr Ala Leu Arg Arg His Ile Ser Tyr Leu Pro  
 595 600 605

55 Gln Gln Ala Tyr Val Phe Ser Gly Ser Ile Met Asp Asn Leu Val Leu

	610	615	620
5	Gly Ala Lys Glu Gly Thr Ser Gln Glu Asp Ile Ile Arg Ala Cys Glu 625 630 635 640		
10	Ile Ala Glu Ile Arg Ser Asp Ile Glu Gln Met Pro Gln Gly Tyr Gln 645 650 655		
15	Thr Glu Leu Ser Asp Gly Ala Gly Ile Ser Gly Gly Gln Lys Gln Arg 660 665 670		
20	Ile Ala Leu Ala Arg Ala Leu Leu Thr Gln Ala Pro Val Leu Ile Leu 675 680 685		
25	Asp Glu Ala Thr Ser Ser Leu Asp Ile Leu Thr Glu Lys Lys Ile Ile 690 695 700		
30	Ser Asn Leu Leu Gln Met Thr Glu Lys Thr Ile Ile Phe Val Ala His 705 710 715 720		
35	Arg Leu Ser Ile Ser Gln Arg Thr Asp Glu Val Ile Val Met Asp Gln 725 730 735		
40	Gly Lys Ile Val Glu Gln Gly Thr His Lys Glu Leu Leu Ala Lys Gln 740 745 750		
45	Gly Phe Tyr Tyr Asn Leu Phe Asn 755 760		
50	<210> 27		
	<211> 900		
	<212> DNA		
55	<213> Streptococcus mutans		
	<220>		
	<221> CDS		
	<222> (1) .. (900)		

<400> 27		
	atg gat cct aaa ttt tta caa agt gca gaa ttt tat agg aga cgc tat	48
	Met Asp Pro Lys Phe Leu Gln Ser Ala Glu Phe Tyr Arg Arg Arg Tyr	
5	1 5 10 15	
	cat aat ttt gcg aca cta tta att gtt cct ttg gtc tgc ttg att atc	96
	His Asn Phe Ala Thr Leu Leu Ile Val Pro Leu Val Cys Leu Ile Ile	
	20 25 30	
10	ttc ttg gtc ata ttc ctt tgt ttt gct aaa aaa gaa att aca gtg att	144
	Phe Leu Val Ile Phe Leu Cys Phe Ala Lys Lys Glu Ile Thr Val Ile	
	35 40 45	
15	tct act ggt gaa gtt gca cca aca aag gtt gta gat gtt atc caa tct	192
	Ser Thr Gly Glu Val Ala Pro Thr Lys Val Val Asp Val Ile Gln Ser	
	50 55 60	
20	tac agt gac agt tca atc att aaa aat aat tta gat aat aat gca gct	240
	Tyr Ser Asp Ser Ser Ile Ile Lys Asn Asn Leu Asp Asn Asn Ala Ala	
	65 70 75 80	
25	gtt gag aag gga gac gtt tta att gaa tat tca gaa aat gcc agt cca	288
	Val Glu Lys Gly Asp Val Leu Ile Glu Tyr Ser Glu Asn Ala Ser Pro	
	85 90 95	
	aac cgt cag act gaa caa aag aat att ata aaa gaa aga caa aaa cga	336
	Asn Arg Gln Thr Glu Gln Lys Asn Ile Ile Lys Glu Arg Gln Lys Arg	
	100 105 110	
30	gaa gag aag gaa aag aaa aaa cac caa aag agc aag aaa aag aag aag	384
	Glu Glu Lys Glu Lys Lys Lys His Gln Lys Ser Lys Lys Lys Lys Lys	
	115 120 125	
35	tct aag agc aag aaa gct tcc aaa gat aag aaa aag aaa tcg aaa gac	432
	Ser Lys Ser Lys Lys Ala Ser Lys Asp Lys Lys Lys Lys Ser Lys Asp	
	130 135 140	
40	aag gaa agc agc tct gac gat gaa aat gag aca aaa aag gtt tcg att	480
	Lys Glu Ser Ser Ser Asp Asp Glu Asn Glu Thr Lys Lys Val Ser Ile	
	145 150 155 160	
45	ttt gct tca gaa gat ggt att att cat acc aat ccc aaa tat gat ggt	528
	Phe Ala Ser Glu Asp Gly Ile Ile His Thr Asn Pro Lys Tyr Asp Gly	
	165 170 175	
	gcc aat att att ccg aag caa acc gag att gct caa atc tat cct gat	576
	Ala Asn Ile Ile Pro Lys Gln Thr Glu Ile Ala Gln Ile Tyr Pro Asp	
	180 185 190	
50	att caa aaa aca aga aaa gtg tta atc acc tat tat gct tct tct gat	624
	Ile Gln Lys Thr Arg Lys Val Leu Ile Thr Tyr Tyr Ala Ser Ser Asp	
	195 200 205	
55	gat gtt gtt tct atg aaa aag ggg caa acc gct cgt ctt tcc ttg gaa	672
	Asp Val Val Ser Met Lys Lys Gly Gln Thr Ala Arg Leu Ser Leu Glu	
	210 215 220	

aaa aag gga aat gac aag gtt gtt att gaa gga aaa att aac aat gtc 720  
 Lys Lys Gly Asn Asp Lys Val Val Ile Glu Gly Lys Ile Asn Asn Val  
 225 230 235 240

5 gct tca tca gca act act act aaa aaa gga aat ctc ttt aag gtt act 768  
 Ala Ser Ser Ala Thr Thr Thr Lys Lys Gly Asn Leu Phe Lys Val Thr  
 245 250 255

10 gcc aaa gta aag gtt tct aag aaa aat agc aaa ctc atc aag tat ggt 816  
 Ala Lys Val Lys Val Ser Lys Lys Asn Ser Lys Leu Ile Lys Tyr Gly  
 260 265 270

15 atg aca ggc aag aca gtc act gtc att gat aaa aag act tat ttt gat 864  
 Met Thr Gly Lys Thr Val Thr Val Ile Asp Lys Lys Thr Tyr Phe Asp  
 275 280 285

20 tat ttc aaa gat aaa tta ctg cat aaa atg gat aat 900  
 Tyr Phe Lys Asp Lys Leu Leu His Lys Met Asp Asn  
 290 295 300

<210> 28

25 <211> 300

<212> PRT

<213> Streptococcus mutans

30 <400> 28

35 Met Asp Pro Lys Phe Leu Gln Ser Ala Glu Phe Tyr Arg Arg Arg Tyr  
 1 5 10 15

40 His Asn Phe Ala Thr Leu Leu Ile Val Pro Leu Val Cys Leu Ile Ile  
 20 25 30

Phe Leu Val Ile Phe Leu Cys Phe Ala Lys Lys Glu Ile Thr Val Ile  
 35 40 45

45 Ser Thr Gly Glu Val Ala Pro Thr Lys Val Val Asp Val Ile Gln Ser  
 50 55 60

50 Tyr Ser Asp Ser Ser Ile Ile Lys Asn Asn Leu Asp Asn Asn Ala Ala  
 65 70 75 80

55 Val Glu Lys Gly Asp Val Leu Ile Glu Tyr Ser Glu Asn Ala Ser Pro  
 85 90 95

Asn Arg Gln Thr Glu Gln Lys Asn Ile Ile Lys Glu Arg Gln Lys Arg  
 100 105 110

5 Glu Glu Lys Glu Lys Lys Lys His Gln Lys Ser Lys Lys Lys Lys  
 115 120 125

10 Ser Lys Ser Lys Lys Ala Ser Lys Asp Lys Lys Lys Lys Ser Lys Asp  
 130 135 140

15 Lys Glu Ser Ser Ser Asp Asp Glu Asn Glu Thr Lys Lys Val Ser Ile  
 145 150 155 160

Phe Ala Ser Glu Asp Gly Ile Ile His Thr Asn Pro Lys Tyr Asp Gly  
 165 170 175

20 Ala Asn Ile Ile Pro Lys Gln Thr Glu Ile Ala Gln Ile Tyr Pro Asp  
 180 185 190

25 Ile Gln Lys Thr Arg Lys Val Leu Ile Thr Tyr Tyr Ala Ser Ser Asp  
 195 200 205

30 Asp Val Val Ser Met Lys Lys Gly Gln Thr Ala Arg Leu Ser Leu Glu  
 210 215 220

35 Lys Lys Gly Asn Asp Lys Val Val Ile Glu Gly Lys Ile Asn Asn Val  
 225 230 235 240

Ala Ser Ser Ala Thr Thr Thr Lys Lys Gly Asn Leu Phe Lys Val Thr  
 245 250 255

40 Ala Lys Val Lys Val Ser Lys Lys Asn Ser Lys Leu Ile Lys Tyr Gly  
 260 265 270

45 Met Thr Gly Lys Thr Val Thr Val Ile Asp Lys Lys Thr Tyr Phe Asp  
 275 280 285

50 Tyr Phe Lys Asp Lys Leu Leu His Lys Met Asp Asn  
 290 295 300

**We claim:**

1. A compound that competitiv ly inhibits binding of CSP to *S. mutans* histidin kinase.
2. The compound of claim 1, comprising a peptide or an antibody.
3. The compound of claim 2, comprising a derivative of [SEQ ID NO:2], a fragment of  
5 [SEQ ID NO:2] or a derivative of a fragment of [SEQ ID NO:2].
4. The compound of claim 3, wherein amino acids are removed from the N-terminus and/or C-terminus of [SEQ ID NO:2].
5. A pharmaceutical composition comprising all or part of the peptide of claims 1 to 4 and a carrier.
- 10 6. A method of medical treatment or prophylaxis of caries or endocarditis, comprising administering the compound of any of claims 1 to 4 or the pharmaceutical composition of claim 5.
7. An isolated nucleic acid molecule encoding a *S. mutans* competence signal peptide, or a fragment of a peptide having CSP activity.
- 15 8. An isolated nucleic acid molecule encoding a competence signal peptide, or a fragment of a competence signal peptide having *S. mutans* competence signal peptide activity, comprising a nucleic acid molecule selected from the group consisting of:
  - (a) a nucleic acid molecule that hybridizes to all or part of a nucleic acid  
20 molecule shown in [SEQ ID NO:1], the fragment of [SEQ. ID NO:1] encoding [SEQ ID NO:2] or a complement thereof under moderate or high stringency hybridization conditions;
  - (b) a nucleic acid molecule degenerate with respect to (a).
- 25 9. An isolated nucleic acid molecule encoding a competence signal peptide, or a fragment of a competence signal peptide having *S. mutans* competence signal peptide activity, comprising a nucleic acid molecule selected from the group consisting of:
  - (a) the nucleic acid molecule of the coding strand shown in [SEQ ID NO:1], or a complement thereof;

- (b) a nucleic acid molecule encoding the same amino acid sequence as a nucleotide sequence of (a); and
- (c) a nucleic acid molecule having at least 50% or 60% identity with the nucleotide sequence of (a) or the fragment of [SEQ. ID NO:1] encoding [SEQ. ID NO:2].

- 5 10. The nucleic acid molecule of claim 9, comprising all or part of a nucleotide sequence shown in [SEQ. ID NO:1], the fragment of [SEQ. ID NO:1] encoding [SEQ. ID NO:2], or a complement thereof.
- 10 11. A CSP nucleic acid molecule isolated from *S. mutans*, or a fragment thereof having CSP activity.
12. A recombinant nucleic acid molecule comprising a nucleic acid molecule of any of claims 7 to 11 and a constitutive promoter sequence or an inducible promoter sequence, operatively linked so that the promoter enhances transcription of the nucleic acid molecule in a host cell.
- 15 13. The nucleic acid molecule of any of claims 7 to 11, wherein the molecule comprises genomic DNA, or cDNA.
14. The nucleic acid molecule of any of claims 7 to 11, wherein the nucleic acid molecule is chemically synthesized.
- 20 15. The nucleic acid molecule of any of claims 7 to 11, wherein the CSP is involved in genetic competence, biofilm formation, and acid tolerance of *S. mutans*.
16. A vector comprising the nucleic acid molecule of any of claims 7 to 11.
17. A host cell comprising the recombinant nucleic acid molecule of claim 12 or the vector of claim 16, or progeny of the host cell.
- 25 18. The host cell of claim 17, selected from the group consisting of a fungal cell, a yeast cell, a bacterial cell, a mammalian cell and a plant cell.
19. A method for producing a recombinant host cell capable of expressing the nucleic acid molecule of any of claims 7 to 11, the method comprising introducing into the host cell a vector of claim 16.
- 30 20. A method for expressing a peptide in the host cell of claim 19, the method comprising culturing the host cell under conditions suitable for gene expression.



21. An isolated polypeptide encoded by and/or produced from the nucleic acid molecule of any of claims 7 to 11, or the vector of claim 16.
22. An isolated CSP or a fragment thereof having *S. mutans* CSP activity.
23. The polypeptide of claim 22 comprising a *S. mutans* CSP.
- 5 24. The polypeptide of claim 23 comprising all or part of an amino acid sequence in [SEQ ID NO:2].
25. A polypeptide fragment of the peptide of claim 24, or a peptide mimetic of the CSP.
26. The polypeptide of claim 24 which is recombinantly produced.
27. A polypeptide comprising a sequence having greater than 30%, 50% or 60%  
10 sequence identity to the polypeptide of claim 24.
28. The polypeptide of claim 24, isolated from *S. mutans*.
29. An isolated nucleic acid molecule encoding the polypeptide of any of claims 21 to 28.
30. An antibody directed against the polypeptide of any of claims 21 to 28.
31. The antibody of claim 30, comprising a monoclonal antibody or a polyclonal antibody.
- 15 32. A vaccine composition comprising all or part of the peptide of any of claims 21 to 28 and a carrier.
33. The vaccine composition of claim 32, wherein the peptide is coupled to a compound comprising all or part of KLH, ovalbumin, or thyroglobulin.
34. A method of evaluating caries-reducing properties of a compound comprising  
20 contacting the compound with:
  - a) CSP, a HK-binding fragment of CSP or a derivative of either of the foregoing; and
  - b) HK, a CSP binding fragment of HK or a derivative of either of the foregoing;
- 25 wherein (a) and (b) are capable of binding; and determining the ability of the compound to interfere with the binding of a) with b), the ability to interfere with binding indicating that the compound reduces caries.
35. A method of evaluating caries-reducing properties of a compound comprising contacting the compound with:

a) a DNA vector encoding a marker gene; and

b) a *S. mutans* culture;

by determining whether the compound reduces uptake of the DNA vector into the *S. mutans* culture, the reduced uptake of the DNA vector indicating that the compound  
5 reduces caries.

36. The method of claim 35, wherein reduction of caries is indicted by reduced transformation efficiency in *S. mutans*.

37. The method of claim 35, wherein reduction of caries is indicted by determining changes in the physiological characteristics of biofilm formation and acid tolerance in *S. mutans*.  
10

Figure 1

Schematic layout of the arrangement of the genetic locus encoding the signal peptide precursor, the histidine kinase and the response regulator. Note that this arrangement is different from other loci in related streptococci for the following reasons: a) The *comC* gene is transcribed from its own promoter alone, unlike the genes thus far described in other streptococci that are arranged in an operon-like cluster with the *comC/DE* genes being transcribed from a single promoter.

b) The *comC* gene is separated from the *comD* gene by 148 nucleotides.

**Streptococcus mutans**  
***ComCDE* Operon**



## Figure 2

Sequences of the open reading frames encoding the signal peptide precursor (ComC), the histidine kinase (ComD), and the response regulator (ComE).

➤ *S. mutans* comC gene

Encodes a precursor to a signal peptide

[ATGAAAAAACACTATCATTAAAAAATGACTTTAAAGAAATTAAGACTGATGAATTAG-  
AGATTATCATTGGCGGA (AGCGGAAGCCTATCAACATTTTTCCGGCTGTTTAACAGAAG  
TTTTACACAAGCTTTGGGAAAA) ] TAA

➤ *S. mutans* CSP encoding sequence

Competence Signal Peptide

AGCGGAAGCCTATCAACATTTTTCCGGCTGTTTAACAGAAGTTTTACACAAGCTTTGGG  
AAAA [SEQ ID NO:1]

➤ *S. mutans* comD gene

Encodes a protein that functions as a histidine kinase  
receptor

[ATGAATGAAGCCTTAATGATACTTTCAAATGGTTTATTAACCTATCTAACCGTTCTAT  
TTCTCTTGTCTTCTATTTTCTAAGGTAAGTAATGTCACCTTTATCGAAAAAGGAATTAAC  
CTTTTTTCGATAAGCAATTTTCTGATAATGATTGCTGTTACGATGGTGAACGTAAACCT  
GTTTTATCCTGCAGAGCCTCTTTATTTTATAGCTTTATCAATTTATCTTAATAGACAGA  
ATAGTCTTTCTCTAAATATATTTTATGGTCTGCTGCCTGTTGCCAGTTCTGACTTGTTT  
AGGCGGGCAATCATATTCTTTATCTTGGATGGAACCTCAAGGAATTGTAATGGGCAGTAG  
CATTATAACCACCTATATGATCGAGTTTGCAGGAATAGCGCTAAGTTACCTCTTTCTCA  
GTGTGTTCAATGTTGATATTGGTCGACTTAAAGATAGTTTGACCAAGATGAAGGTCAAA  
AAACGCTTGATTCCAATGAATATTACTATGCTTCTATACTACCTTTTAATACAGGTATT  
GTATGTTATAGAGAGTTATAATGTGATACCGACTTTAAATTTTCGTAAATTTGTCGTTA  
TTGTCTATCTTATTTTATTTTGTATTCTGATCTCATTTTTAAGCCAATATACCAAACAA  
AAGGTTCAAATAGATAATGGCACAAAAGGAAGCTCAGATTGAAATATCACCCAGTA  
TAGTCAGCAAATAGAATCTCTTTACAAGGATATTGGAAGTTTCCGCCATGATTATCTGA  
ATATTTTAACTAGCCTCAGATTAGGCATTGAAAATAAAGATTTAGCTAGTATTGAAAAG  
ATTTACCATCAAATCTTAGAAAAAACAGGACATCAATTGCAGGATACCCGTTATAATAT  
CGGCCATCTAGCTAATATTCAAACGATGCTGTCAAGGGTATCTTGTGAGCAAAAATCT  
TAGAAGCTCAGAATAAAAAGATTGCTGTCAATGTAGAAGTCTCAAGTAAAATACAACCTG  
CCTGAGATGGAGTTGCTTGATTTTATTACCATACTTTCTATCTTGTGTGATAATGCCAT  
TGAGGCTGCTTTTGAATCATTAAATCCTGAAATTCAGTTAGCCTTTTTTAAGAAAAATG  
GCAGTATAGTCTTTATCATTGAGAATTCACCAAAGAAAAACAAATAGATGTGAGTAAA  
ATTTTTAAAGAACTATTCCACTAAAGGCTCCAATCGCGGTATTGGTTTAGCAAAGGTG  
AATCATATTCTTGAACATTATCCCAAACAGTTTACAAACAAGCAATCATCATCATTT  
ATTCAAGCAACTCCTAATAATAAAA] TAG

➤ *S. mutans* comE gene

Encodes a response regulator that activates transcription  
of a number of genes

[ATGATTTCTATTTTTGTATTGGAAGATGATTTTTTACAACAAGGACGTCTTGAAACCA  
CCATTGCAGCTATCATGAAAGAAAAAATTGGTCTTATAAAGAATTGACTATTTTTGGA  
AAACCACAACAACTTATTGACGCTATCCCTGAAAAGGGCAATCACCAGATTTTCTTTT  
GGATATTGAAATCAAAAAAGAGGAAAAGAAAGGACTGGAAGTAGCCAATCAGATTAGAC  
AGCATAATCCTAGTGCAGTTATTGTCTTTGTCACGACACATTCTGAGTTTATGCCCCCTC  
ACTTTTCAGTATCAGGTATCTGCTTTGGATTTTATTGATAAATCTTTGAATCCTGAGGA  
GTTCTCCCACCGCATTGAATCAGCGCTGTATTATGCTATGGAAAACAGCCAGAAGAATG  
GTCAATCAGAGGAACTTTTTATTTTCCATTCTGAAACTCAGTTTCAGGTCCCTTTT  
GCTGAGATTCTGTATTTTGAACATCTTCAACAGCCCATAAGCTCTGCCTTTATACTTA  
TGATGAACGGATTGAATTCTACGGCAGTATGACTGACATTGTTAAAATGGATAAGAGAC  
TTTTTTCAGTGCCATCGCTCTTTTATTGTCAATCCTGCCAATATTACCCGTATTGATCGG  
AAAAAACGCTTGGCCTATTTTCGAAATAATAAGTCTTGTCTTATTTACGAACTAAGTT  
AACAAAACGAGAGCTGTGATTGCTGATCAAAGGAGAGCAAAA] TGA

## Figure 3

The amino acid sequences of the signal peptide precursor (ComC), the histidine kinase (ComD), and the response regulator (ComE).

➤ *S. mutans* ComC protein (CSP Precursor)

MKKTLSLKNDFKEIKTDELEIIIGSGSLSTFFRLFNRSFTQALGK

➤ *S. mutans* ComD protein (Histidine Kinase)

MNEALMILSNGLLTLYLTVLFLFLFSKVSNTLSKKELTFSISNFLIMIAVTMVNVNL  
FYPAEPLYFIALSIYLNQRNSLSLNIFYGLLPVASSDLFERRAIIFFILDGTQGIVMGSS  
IITTYMIEFAGIALSYLFLSVFNVDIGRLKDSLTKMKVKKRLIPMNITMLLYYLLIQVL  
YVIESYNVIPTLKFRKFVIVYLILFLILISFLSQYTKQKVQNEIMAQKEAQIRNITQY  
SQQIESLYKDIRSFRHDYLNILTSRLGIENKDLASIEKIYHQILEKTGHQLQDTRYNI  
GHLANIQNDAVKGILSAKILEAQNKKIADVNEVSSKIQLPEMELLDFFITLSILCDNAI  
EAAFESLNPEIQLAFFKKNGSIVFIIQNSTKEKQIDVSKIFKENYSTKGSNRGIGLAKV  
NHILEHYPKTSLQTSNHHHLFKQLLIK

➤ *S. mutans* ComE protein (Response Regulator)

MISIFVLEDDFLQQGRLETTIAAIMKEKNWSYKELTIFGKPQQQLIDAIPEKGNHQIFFL  
DIEIKKEEKKGLEVANQIRQHNPSAVIVFVTTTHSEFMPLTFQYQVSALDFIDKSLNPEE  
FSHRIESALYYAMENSQKNGQSEELFIFHSSETQFQVPPAEILYFETSSTAHLCLYTY  
DERIEFYGSMTDIVKMDKRLFQCHRSFIVNPANITRIDRKKRLAYFRNNKSCLISRTKL  
TKLRAVIADQRRAK

## Figure 4

The deduced amino acid sequence of the signal peptide precursor in various strains and its predicted cleavage site. The original peptide is expressed as a 46-amino acid peptide that is cleaved after the glycine-glycine residues to generate an active signal peptide.

```

BM71 CSP      1 MKKTPSLKNDKFKEIKTDELEIIIGGSGSLSTFFRLFNRSFTQALGK 46
GB14 CSP      1 MKKTLSLKNDKFKEIKTDELEIIIGGSGSLSTFFRLFNRSFTQALGK 46
H7 CSP        1 MKKTLSLKNDKFKEIKTDELEIIIGGSGSLSTFFRLFNRSFTQALGK 46
JH1005 CSP    1 MKKTLSLKNDKFKEIKTDELEIIIGGSGTLSTFFRLFNRSFTQA   43
LT11 CSP      1 MKKTLSLKNDKFKEIKTDELEIIIGGSGSLSTFFRLFNRSFTQALGK 46
NG8 CSP       1 MKKTLSLKNDKFKEIKTDELEIIIGGSGSLSTFFRLFNRSFTQALGK 46
UAB159 CSP    1 MKKTLSLKNDKFKEIKTDELEIIIGGSGSLSTFFRLFNRSFTQALGK 46
      **** *

```

```

consensus: 1 MKKTLSLKNDKFKEIKTDELEIIIGG SGSLSTFFRLFNRSFTQALGK 46
predicted cleavage site:                ^

```

## Figure 5

The synthetic signal peptide that is effective at inducing competence, biofilm formation and acid tolerance in *Streptococcus mutans*.

SGSLSTFFRLFNRSFTQALGK [SEQ ID NO:2]

**Figure 6**

The natural activity of the signal/receptor system functioning *in vitro* in model biofilms as determined by the ability of various strains of *S. mutans* to accept donor plasmid DNA conferring erythromycin resistance.

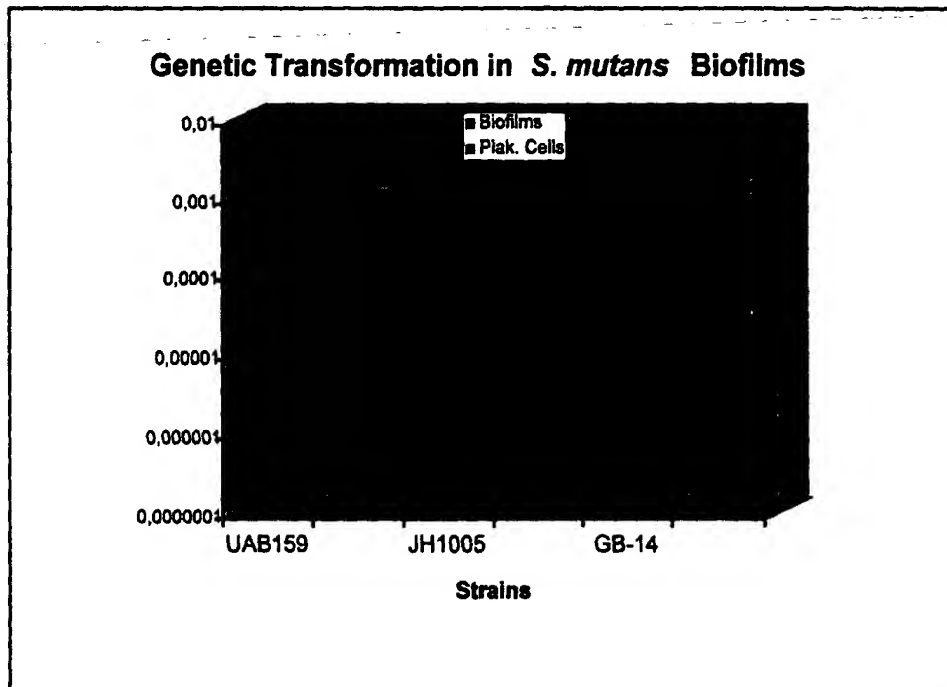




Figure 7  
Induction of genetic transformation in *Streptococcus mutans*  
by synthetic competence stimulating peptide (SCSP)<sup>1</sup>

Strain	Peptide added Number of Transformants/Recipients	No peptide Number of Transformants/Recipients
UAB15	$4.65 \times 10^{-1}$	$1.78 \times 10^{-6}$
JH1005 <sup>2</sup>	$6.98 \times 10^{-2}$	0

<sup>1</sup>The final concentration of SCSP used was 500 ng/ml.

The strain contains a nonsense mutation in the *comC* gene encoding the CSP.

## Figure 8

List of the primers used to amplify the genes or internal regions of the target genes by polymerase chain reaction (PCR) for subsequent sequencing or inactivation.

## ComC region

## ComC Primer Pair: F5-B5

- [F5] 23406-23424 5'- AGTTTTTTGTCTGGCTGCG -3'  
 19 nt forward primer  
 pct G+C: 47.4 Tm: 50.5
- [B5] 24056-24037 5'- TCCACTAAAGGCTCCAATCG -3'  
 20 nt backward primer  
 pct G+C: 50.0 Tm: 51.9
- 651 nt product for F5-B5 pair (23406-24056)  
 Optimal annealing temp: 50.3  
 pct G+C: 30.9 Tm: 71.5

## ComD region

## ComD Primer Pair: F1-B1

- [F1] 392-415 5'- CGCTAAGTTACCTCTTTCTCAGTG -3'  
 24 nt forward primer  
 pct G+C: 45.8 Tm: 51.6
- [B1] 683-663 5'- GCTTCCTTTTGTGCCATTATC -3'  
 21 nt backward primer  
 pct G+C: 42.9 Tm: 50.8
- 292 nt product for F1-B1 pair (392-683)  
 Optimal annealing temp: 49.5  
 pct G+C: 30.8 Tm: 70.2

## ComE region

## ComE Primer Pair: F1-B1

- [F1] 145-165 5'- CCTGAAAAGGGCAATCACCAG -3'  
 21 nt forward primer  
 pct G+C: 52.4 Tm: 55.9
- [B1] 606-585 5'- GCGATGGCACTGAAAAAGTCTC -3'  
 22 nt backward primer  
 pct G+C: 50.0 Tm: 55.4
- 462 nt product for F1-B1 pair (145-606)  
 Optimal annealing temp: 53.6  
 pct G+C: 38.3 Tm: 74.1

Figure 9

ComCDE local region. The ComC (first highlighted region; nucleotides 101 to 241), ComD (second highlighted region; nucleotides 383 to 1708) and ComE (third highlighted region; nucleotides 1705 to 2457) proteins are highlighted.

Sequence Range: 1 to 2557

```

      10      20      30      40      50
ACATTATGTGTCCTAAGGAAAATATTACTTTTTCAAGAAAATCCATGATT
TGTAATACACAGGATTCCTTTTATAATGAAAAAGTTCTTTTAGGTACTAA
      <K K L F I W S K
      <
      60      70      80      90     100
TTTTCATAAAAAATAGTATACTAATTATAATCAAAAAAGGAGATATAAA
AAAAGTATTTTTATCATATGATTAATATTAGTTTTTTCTCTATATTT
      <K M F F L I S I I I L F L L Y L
      <
      110     120     130     140     150
ATGAAAAAACACTATCATTAAAAATGACTTTAAAGAAATTAAGACTGA
TACTTTTTTGTGATAGTAATTTTTTACTGAAATTTCTTTAATTCTGACT
      M K K T L S L K N D P K E I K T D >
      ORF RF[2]
      <I F F V S D N F F S K L S I L V S
      <
      160     170     180     190     200
TGAATTAGAGATTATCATTGGCGGAAGCGGAAGCCTATCAACATTTTTCC
ACTTAATCTCTAATAGTAACCGCCTTCGCCTTCGGATAGTTGTAAAAAGG
      E L E I I I G G S G S L S T F F >
      ORF RF[2]
      <S N S I I M
      <
      210     220     230     240     250
GGCTGTTTAACAGAAGTTTTACACAAGCTTTGGGAAAATAAGATAGGCTA
CCGACAAATGTCTTCAAAATGTGTTCGAAACCCTTTTATTCTATCCGAT
      R L F N R S F T Q A L G K >
      ORF RF[2]
      <
      260     270     280     290     300
ACATTGGAATAAAACAAGGCTGGATTATTATTCCAGCCTTTTAAATGT
TGTAACCTTATTTTGTCCGACCTAAATAATAAGGTCCGAAAAATTACA
      310     320     330     340     350
AAAATAAAAAATACAGGGTTAAATAATCAAGTGTGCTGTCTGGATGAGAA
TTTTATTTTATGTCCAATTTATTAGTTCACACGACAGCACCTACTCTT
      360     370     380     390     400
GATAAACTATCTCTTAGAGAATAGGCCTCCTCTATTTTATTATTAGGAG
CTATTTTGATAGAGAATCTCTTATCCGGAGGAGATAAAATAATAATCCTC
      <K I I L L
      < ORF RF[
      410     420     430     440     450

```

TTGCTTGAATAAATGATGATGATTGCTTGTGTTGTAACCTGGTTTTGGGAT  
 AACGAACCTTATTTACTACTACTAACGAACAAACATTTGACCAAAACCCTA  
 <Q K F L H H N S T Q L S T K P Y  
 < ORF RF[4] C

460 470 480 490 500  
 AATGTTCAAGAATATGATTACCTTTGCTAAACCAATACCGGATTGGAG  
 TTACAAGTTCTTATACTAAGTGGAACGATTTGGTTATGGCGCTAACCTC  
 <H E L I H N V K A L G I G R N S  
 < ORF RF[4] C

510 520 530 540 550  
 CCTTTAGTGAATAGTTTTCTTTAAAAATTTTACTCACATCTATTTGTTT  
 GGAAATCACCTTATCAAAGAAATTTTAAAATGAGTGTAGATAAACAAA  
 <G K T S Y N E K F I K S V D I Q K  
 < ORF RF[4] C

560 570 580 590 600  
 TTCTTTGGTGAATTCTGAATGATAAAGACTATACTGCCATTTTCTTAA  
 AAGAAACCACTTAAGACTTACTATTCTGATATGACGGTAAAAAGAATT  
 <E K T S N Q I I F V I S G N K R F  
 < ORF RF[4] C

610 620 630 640 650  
 AAAAGGCTAACTGAATTTCAGGATTTAATGATTGAAAGCAGCCTCAATG  
 TTTTCCGATTGACTTAAAGTCCTAAATTACTAAGCTTTCGTCCGAGTTAC  
 M>

<P A L Q I E P N L S E F A A E I  
 < ORF RF[4] C

660 670 680 690 700  
 GCATTATCACACAAGATAGAAAGTATGGTAATGAAATCAAGCAACTCCAT  
 CGTAATAGTGTGTTCTATCTTTCATACCATTACTTTAGTTTCGTTGAGGTA  
 A L S H K I E S M V M K S S N S I>  
 ORF RF[3]  
 <A N D C L I S L I T I F D L L E M  
 < ORF RF[4] C

710 720 730 740 750  
 CTCAGGCAGTTGTATTTTACTTGAGACTTCTACATTGACAGCAATCTTTT  
 GAGTCCGTCAACATAAAATGAACTCTGAAGATGTAAGTGTCTGTTAGAAAA  
 S G S C I L L E T S T L T A I F>  
 ORF RF[3]  
 <E P L Q I K S S V E V N V A I K K  
 < ORF RF[4] C

760 770 780 790 800  
 TATCTGAGCTTCTAAGATTTTGTGACAAGATACCCTTGACAGCATCG  
 ATAAGACTCGAAGATTCTAAAAACGACTGTTCTATGGGAAGTGTCTAGC  
 L F>

<N Q A E L I K A S L I G K V A D  
 < ORF RF[4] C

810 820 830 840 850  
 TTTTGAATATTAGCTAGATGGCCGATATTATAACGGGTATCCTGCAATTG

AAAAC TTATAATCGATCTACCGGCTATAATATTGCCCATAGGACGTTAAC  
 <N Q I N A L H G I N Y R T D Q L Q  
 < ORF RF[4] C

860 870 880 890 900  
 ATGTCCTGTTTTCTAAGATTTGATGGTAAATCTTTCAATACTAGCTA  
 TACAGGACAAAAAGATTCTAAACTACCATTAGAAAAAGTTATGATCGAT  
 <H G T K E L I Q H Y I K E I S A L  
 < ORF RF[4] C

910 920 930 940 950  
 AATCTTTATTTCAATGCCTAATCTGAGGCTAGTTAAAATATTGAGATAA  
 TTAGAAATAAAAGTTACGGATTAGACTCCGATCAATTTTATAAGTCTATT  
 <D K N E I G L R L S T L I N L Y  
 < ORF RF[4] C

960 970 980 990 1000  
 TCATGGCGGAAACTTCGAATATCCTTGTAAGAGATTCTATTTGCTGACT  
 AGTACCGCCTTTGAAGCTTATAGGAACATTTCTCTAAGATAAACGACTGA  
 M A E T S N I L V K R F Y L L T>

<D H R F S R I D K Y L S E T Q Q S  
 < ORF RF[4] C

1010 1020 1030 1040 1050  
 ATACTGGGTGATATTTTGAATCTGAGCTTCCTTTGTGCCATTATCTCAT  
 TATGACCCACTATAAAGCTTAGACTCGAAGGAAAACACGGTAATAGAGTA  
 I L G D I S N L S F L L C H Y L I>

<Y Q T I N R I Q A E K Q A M I E N  
 < ORF RF[4] C

1060 1070 1080 1090 1100  
 TTTGAACCTTTTGTGTTGTTATATTGGCTTAAAAATGAGATCAGAATCAAA  
 AAACCTTGAAAAACAAACCATATAACCGAATTTTACTCTAGTCTTAGTTT  
 L N L L F G I L A>

<Q V K Q K T Y Q S L F S I L S L  
 < ORF RF[4] C

1110 1120 1130 1140 1150  
 AATAAAATAAGATAGACAATAACGACAAATTTACGAAATTTTAAAGTCGG  
 TTATTTTATTCTATCTGTTATTGCTGTTAAATGCTTTAAATTTTCAGCC  
 <F L I L Y W I V V F K R F K L T P  
 < ORF RF[4] C

1160 1170 1180 1190 1200  
 TATCACATTATAACTCTCTATAACATACAATACCTGTATTAAAAGGTAGT  
 ATAGTGTAATATTGAGAGATATTGTATGTTATGGACATAATTTCCATCA  
 <I V N Y S E T V Y L V Q I L L Y Y  
 < ORF RF[4] C

1210 1220 1230 1240 1250  
 ATAGAAGCATAGTAATATTTCATTGGAATCAAGCGTTTTTTGACCTTCATC  
 TATCTTCGTATCATTATAAGTAACCTTAGTTCGAAAAAAGTGAAGTAG  
 <L L M T I N M P I L R K K V K M  
 < ORF RF[4] C

1260 1270 1280 1290 1300  
 TTGGTCAAACATCTTTAAGTCGACCAATATCAACATTGAACACACTGAG  
 AACCAGTTTGATAGAAATTCAGCTGGTTATAGTTGTAACCTGTGTGACTC  
 <K T L S D K L R G I D V N P V S L  
 < ORF RF[4] C

1310 1320 1330 1340 1350  
 AAAGAGGTAACCTAGCGCTATTCTGCAAACTCGATCATATAGGTGGTTA  
 TTTCTCCATTGAATCGCGATAAGGACGTTTGAGCTAGTATATCCACCAAT  
 <F L Y S L A I G A F E I M Y T T I  
 < ORF RF[4] C

1360 1370 1380 1390 1400  
 TAATGCTACTGCCCATTAACAATTCCTTGAGTTCCATCCAAGATAAAGAAT  
 ATTACGATGACGGGTAATGTTAAGGAACTCAAGGTAGGTTCTATTTCTTA  
 <I S S G M V Y G Q T G D L I F F  
 < ORF RF[4] C  
 <L E K L E M W S L S Y  
 <

1410 1420 1430 1440 1450  
 ATGATTGCCCGCCTAAACAAGTCAGAACTGGCAACAGGCAGCAGACCATA  
 TACTAACGGGCGGATTGTTTCAGTCTTGACCGTTGTCGCTCGTCTGGTAT  
 <I T A R R E D D S S A V P L L G Y  
 < ORF RF[4] C  
 <S Q G G L C T L V P L L C C V M  
 <

1460 1470 1480 1490 1500  
 AAATATATTTAGAGAAAGACTATTCTGTCTATTAAGATAAATTGATAAAG  
 TTTATATAAATCTCTTTCTGATAAGACAGATAATTCTATTAACTATTTC  
 <F I N L S L S N Q R N L Y I S L A  
 < ORF RF[4] C

1510 1520 1530 1540 1550  
 CTATAAAATAAAGAGGCTCTGCAGGATAAAACAGGTTTACGTTCCACCATC  
 GATATTTTATTCTCCGAGACGTCCTATTTTGTCAAATGCAAGTGGTAG  
 <I F Y L P E A E Y F L N V N V M  
 < ORF RF[4] C

1560 1570 1580 1590 1600  
 GTAACAGCAATCATTATCAGAAAATTGCTTATCGAAAAAGAGTTAATTC  
 CATGTGCGTTAGTAATAGTCTTTTAACGAATAGCTTTTCTCAATTAAG  
 <T V A I M L L F N S I S F L T L E  
 < ORF RF[4] C

1610 1620 1630 1640 1650  
 CTTTTTCGATAAAGTGACATTACTTACCTTAGAAAATAGAAACAAGAGAA  
 GAAAAAGCTATTTCACGTGAATGAATGAATCTTTTATCTTTGTTCTCTT  
 <K K S L T V N S V K S F L F L L F  
 < ORF RF[4] C

1660 1670 1680 1690 1700  
 ATAGAACGGTTAGATAAGTTAATAAACCATTTGAAAGTATCATTAAAGGCT  
 TATCTTGCCAACTATTCAATTATTTGGTAACTTTCATAGTAATTCCGA  
 <L V T L Y T L L G N S L I M L A

< \_\_\_\_\_ ORF RF[4] C \_\_\_\_\_

1710 1720 1730 1740 1750  
TCATTTCATTTTGCTCTCCTTTGATCAGCAATCACAGCTCTCAGTTTGTT  
AGTAAGTAAACGAGAGGAACTAGTCGTTAGTGTGAGAGTCAAAACAA  
<E N M

< \_\_\_\_\_  
<K A R R Q D A I L V A R L K T  
< \_\_\_\_\_ ORF RF[5] C \_\_\_\_\_

1760 1770 1780 1790 1800  
AACTTAGTTCGTGAAATAAGACAAGACTTATTATTTTCGAAAATAGGCCAA  
TTGAATCAAGCACTTTATTCTGTTCTGAATAATAAGCTTTTATCCGGTT  
<L K P R S L L C S K N N R F Y A L  
< \_\_\_\_\_ ORF RF[5] C \_\_\_\_\_

1810 1820 1830 1840 1850  
GCGTTTTTTCGGATCAATACGGGTAATATTGGCAGGATTGACAATAAAAG  
CGCAAAAAGGCTAGTTATGCCATTATAACCGTCCTAACTGTTATTTTC  
<R K K R D L R T I N A P N V I T S  
< \_\_\_\_\_ ORF RF[5] C \_\_\_\_\_

1860 1870 1880 1890 1900  
AGCGATGGCACTGAAAAAGTCTCTTATCCATTTTAACAATGTCAGTCATA  
TCGCTACCGTGACTTTTTCAGAGAATAGGTAATAATTGTTACAGTCAGTAT  
M A L K K S L I H F N N V S H>

ORF RF[1] \_\_\_\_\_>  
<R H C Q F L R K D M K V I D T M  
< \_\_\_\_\_ ORF RF[5] C \_\_\_\_\_

<V

<\_

1910 1920 1930 1940 1950  
CTGCCGTAGAATTCAATCCGTTTCATCATAAGTATAAAGGCAGAGCTTATG  
GACGGCATCTTAAGTTAGGCAAGTAGTATTCATATTTCCGTCCTCGAATAC  
T A V E F N P F I I S I K A E L M>  
ORF RF[1] \_\_\_\_\_>

<S G Y F E F E E D Y T Y L C L K H  
< \_\_\_\_\_ ORF RF[5] C \_\_\_\_\_  
<A T S N L G N M M L I F A S S I  
< \_\_\_\_\_ ORF RF[6] C \_\_\_\_\_

1960 1970 1980 1990 2000  
GGCTGTTGAAGATGTTTCAAATAACAGAATCTCAGCAAAAGGGACCTGAA  
CCGACAACTTCTACAAAGTTTATGTCTTAGAGTCGTTTCCCTGGACTT  
G C>

<A T S S T E F V L I E A F P V Q F  
< \_\_\_\_\_ ORF RF[5] C \_\_\_\_\_  
<P Q Q L H K L I C F R L L L S R F  
< \_\_\_\_\_ ORF RF[6] C \_\_\_\_\_

2010 2020 2030 2040 2050  
ACTGAGTTTCAGATGAATGGAAAATAAAAAGTTCCTCTGATTGACCATTTC  
TGACTCAAAGTCTACTTACCTTTTATTTTCAAGGAGACTAACTGGTAAG  
<Q T E S S H F I F L E E S Q G N  
< \_\_\_\_\_ ORF RF[5] C \_\_\_\_\_

<S L K L H I S F L F N R Q N V M R  
 <\_\_\_\_\_ORF RF[6] C\_\_\_\_\_

2060 2070 2080 2090 2100  
 TTCTGGCTGTTTTCCATAGCATAATACAGCGCTGATTCAATGCGGTGGGA  
 AAGACCGACAAAAGGTATCGTATTATGTCGCGACTAAGTTACGCCACCCT  
 <K Q S N E M A Y Y L A S E I R H S  
 <\_\_\_\_\_ORF RF[5] C\_\_\_\_\_

<R A T K W L M  
 <\_\_\_\_\_ORF RF[6] C\_\_\_\_\_

2110 2120 2130 2140 2150  
 GAACTCCTCAGGATTCAAAGATTTATCAATAAAATCCAAAGCAGATACCT  
 CTTGAGGAGTCCTAAGTTTCTAAATAGTTATTTTAGGTTTCGTCTATGGA  
 <P E E P N L S K D I F D L A S V Q  
 <\_\_\_\_\_ORF RF[5] C\_\_\_\_\_

2160 2170 2180 2190 2200  
 GATACTGAAAAGTGAGGGGCATAAACTCAGAATGTGTCGTGACAAAGACA  
 CTATGACTTTTCACTCCCGTATTTGAGTCTTACACAGCACTGTTTCTGT  
 M C R D K D>

<Y Q P T L R M F E S R T T V F V  
 <\_\_\_\_\_ORF RF[5] C\_\_\_\_\_

2210 2220 2230 2240 2250  
 ATAAGTCACTAGGATTATGCTGTCTAATCTGATTGGCTACTTCCAGTCC  
 TATTGACGTGATCCTAATACGACAGATTAGACTAACCAGTGAAGGTCAGG  
 N N C T R I M L S N L I G Y F Q S>

<I V A S P N H L Q R I Q N A V E L G  
 <\_\_\_\_\_ORF RF[5] C\_\_\_\_\_

2260 2270 2280 2290 2300  
 TTTCTTTTCCTCTTTTTTGATTTCAATATCCAAAAGAAAATCTGGTGAT  
 AAAGAAAAGGAGAAAAAACTAAAGTTATAGGTTTTCTTTTAGACCACTA  
 F L F L F F D F N I Q K E N L V I>

<K K E E K K I E I D L F F I Q H N  
 <\_\_\_\_\_ORF RF[5] C\_\_\_\_\_

2310 2320 2330 2340 2350  
 TGCCCTTTTCAGGGATAGCGTCAATAAGTTGTTGTGGTTTTCCAAAATA  
 ACGGAAAAGTCCCTATCGCAGTTATTCAACAACACCAAAAGGTTTTTAT  
 A L F R D S V N K L L W F S K N>

<G K E P I A D D L Q Q P K G F I  
 <\_\_\_\_\_ORF RF[5] C\_\_\_\_\_

2360 2370 2380 2390 2400  
 GTCAATTCTTTATAAGACCAATTTTTTCTTTTCATGATAGCTGCAATGGT  
 CAGTTAAGAAATATTCTGGTTAAAAAAGAAAGTACTATCGACGTTACCA  
 S Q F F I R P I F F F H D S C N G>

M I A A M V>

<T L E K Y S W N K E K M I A A L T



< \_\_\_\_\_ ORF RF[5] C \_\_\_\_\_  
2410 2420 2430 2440 2450  
GGTTTCAAGACGTCCTTGTGTAATAAATCATCTTCCAATACAAAAATAG  
CCAAAGTTCTGCAGGAACAACATTTTTTAGTAGAAGGTTATGTTTTTATC  
G F K T S L L>

>  
V S R R P C C K K S S S N T K I>  
< T E L R G Q Q L F D D E L V F I S  
< \_\_\_\_\_ ORF RF[5] C \_\_\_\_\_

2460 2470 2480 2490 2500  
AAATCATTATTTCTCCTTTAATCTTCTATTTAGGTTAGCTGATTAACACT  
TTTAGTAATAAAGAGGAAATTAGAAGATAAATCCAATCGACTAATTGTGA  
E I I I S P L I F Y L G>

< I M  
< \_\_\_\_\_

2510 2520 2530 2540 2550  
ATACACAGAAAAGGTATAAAACGATATCACTCAATAAAATCTACTAACTT  
TATGTGTCTTTTCCATATTTTGCTATAGTGAGTTATTTTAGATGATTGAA

AATAACC  
TTATTGG

Figure 10

The comX nucleotide sequence, amino acid sequence, and its local region with 100bp included both upstream and downstream (promoter is upstream).

➤ *S. mutans* comX gene

ATGGAAGAAGATTTTGAAATTGTTTTTAATAAGGTAAAGCCAATTGTATGGAAATTAAG  
CCGTTATTACTTTTATTAAATGTGGACTCGTGAAGATTGGCAACAAGAGGGAATGTTGA  
TTTTGCACCAATTATTAAGGGAACATCCAGAATTAGAAGAGGATGATACAAAATTGTAT  
ATCTATTTTAAGACACGTTTTTCTAATTACATTAAAGATGTTTTGCGTCAGCAAGAAAG  
TCAGAAACGTCGTTTTAATAGAATGTCTTATGAAGAAGTCGGTGAGATTGAACACTGTT  
TGTC AAGTGGCGGTATGCAATTGGATGAATATATTTTATTTTCGTGATAGTTTGCTTGCA  
TATAACAAGGTCTGAGTACTGAAAAGCAAGAGCTGTTTGAGCGCTTGGTAGCAGGAGA  
GCACTTTTTGGGAAGGCAAAGTATGCTGAAAGATTTACGTAAAAAATTAAGTGATTTTA  
AGGAAAAA

➤ *S. mutans* ComX protein

MEEDFEIVFNKVKPIVWKL SRYFIKMWTR EDWQQEGMLILHQLLREHPELEEDDTKLY  
IYFKTRFSNYIKDVL RQQESQKRRFN RMSYEEVGEIEHCLSSGGMQLDEYILFRD SLLA  
YKQGLSTEKQELFERLVAGEHFLGRQ SMLKDLRKKLSDFKEK

➤ *S. mutans* comX gene local region

GTAAATAAAACAGCCAGTTAAGATGGGACATTTATGTCCTGTTCTTAAAGTCTTTTTTCG  
TTTTATAATAATTTTATTATAAAAGGAGGTCATCGTAATAGATGGAAGAAGATTTTGAA  
ATTGTTTTTAATAAGGTAAAGCCAATTGTATGGAAATTAAGCCGTTATTACTTTATTAA  
AATGTGGACTCGTGAAGATTGGCAACAAGAGGGAATGTTGATTTTGCACCAATTATTAA  
GGGAACATCCAGAATTAGAAGAGGATGATACAAAATTGTATATCTATTTTAAGACACGT  
TTTTCTAATTACATTAAAGATGTTTTGCGTCAGCAAGAAAGTCAGAAACGTCGTTTTAA  
TAGAATGTCTTATGAAGAAGTCGGTGAGATTGAACACTGTTTGTCAAGTGGCGGTATGC  
AATTGGATGAATATATTTTATTTTCGTGATAGTTTGCTTGCAATATAACAAGGTCTGAGT  
ACTGAAAAGCAAGAGCTGTTTGAGCGCTTGGTAGCAGGAGAGCACTTTTTGGGAAGGCA  
AAGTATGCTGAAAGATTTACGTAAAAAATTAAGTGATTTTAAGGAAAAATAGTTAAAAA  
GGGAAAGAATGGAACATGTGATTGTACCATTCTTTTTGGTTGAAAATTAAGAAAAGTTA  
TTATAAATTATTGGTTTAACATGCCATATTA

Figure 11

The *comA* and *comB* nucleotide and amino acid sequences. *ComA* and *ComB* are the components of the CSP exporter.

➤ *S. mutans comA* gene

ATGAAACAAGTTATTTATGTTGTTTTAATCGTCATAGCCGTTAACATTCTCTTAGAGAT  
TATCAAAAGAGTAACAAAAAGGGGAGGGACAGTTTCGTCATCTAATCCTTTACCAGATG  
GGCAGTCTAAGTTGTTTTGGCGCAGACATTATAAGCTAGTACCTCAGATTGATACCAGA  
GACTGTGGGCCGGCAGTGCTGGCATCTGTTGCAAAGCATTACGGATCTAATTACTCTAT  
CGCTTATCTGCGGGAACCTCTCAAAGACTAACAAAGCAGGGAACAACAGCTCTTGGCATTG  
TTGAAGCTGCTAAAAAGTTAGGCTTTGAAACACGCTCTATCAAGGCGGATATGACGCTT  
TTTGATTATAATGATTTGACCTATCCTTTTATCGTCCATGTGATTAAAGGAAAACGCTCT  
GCAGCATTATTATGTCGTCTATGGCAGCCAGAATAATCAGCTGATTATTGGAGATCCTG  
ATCCTTCAGTTAAGGTGACTAGGATGAGTAAGGAACGCTTTCAATCAGAGTGGACAGGC  
CTTGCAATTTTCTAGCTCCTCAGCCTAACTATAAGCCTCATAAAGGTGAAAAAATGG  
TTTGTCTAATTTCTTCCCGTTGATCTTTAAGCAGAAAGCTTTGATGACTTATATTATCA  
TAGCTAGCTTGATTGTGACGCTCATTGATATTGTGCGATCATACTATCTCCAAGGAATA  
TTGGACGAGTACATTCTGATCAGCTGATTTCACTTTAGGAATGATTACGATTGGTCT  
GATAATAACCTATATTATCCAGCAGGTGATGGCTTTTGCAAAGAATACCTCTTGGCCG  
TACTCAGTTTGCCTTTAGTCATTGATGTTATCCTGTCTTATATCAACATATTTTTACG  
CTTCCTATGTCTTTCTTTGCGACAAGGCGAACAGGAGAAATCACGTCTCGTTTTACAGA  
TGCCAATCAGATTATTGATGCTGTAGCGTCAACCATCTTTTCAATCTTTTATGATATGA  
CTATGGTAATTTTGGTTGGTGGGGTTTGTGTCGCGCAAAACAATAACCTTTTCTTTCTA  
ACCTTGCTCTCCATTCCGATTTATGCCATCATTATTTTTGCTTTCTTGAAACCCTTTGA  
GAAATGAATCACGAAGTGATGGAAAGCAATGCTGTGGTAAGTTCTTCTATCATTGAAG  
ATATCAATGGGATGGAAACCATTAATCACTCACAAAGTGAGTCCGCTCGTTATCAAAAC  
ATTGATAGTGAATTTGTTGATTATTTGGAGAAAACTTTAAGCTACACAAGTATAGTGC  
CATTCAAACCGCATTAAAAAGCGGTGCTAAGCTTATCCTCAATGTTGTCATTCTCTGGT  
ATGGCTCTCGTCTAGTTATGGATAATAAATCTCAGTTGGTCAGCTTATCACCTTTAAT  
GCTTTGCTGTCTTATTTCTCAAATCCAATTGAAAATATTATCAATCTGCAATCCAACT  
GCAGTCAGCTCGCGTTGCCAATACACGCTTAAATGAGGTCTATCTTGTGCAATCTGAAT  
TTGAAAAAGACGGCGATTTATCAGAAAATAGCTTTTATAGATGGTGATATTTGTTTTGAA  
AATCTTTCTTATAAATATGGATTGGGCGAGATACCTTATCAGATATTAATTTATCAAT  
CAAAAAAGGCTCCAAGGTGAGTCTAGTTGGAGCCAGTGGTTCTGGTAAAACAACCTTTGG  
CTAAACTGATTGTCAATTTCTACGAGCCTAACAAAGGGGATTGTTGCAATCAATGGCAAT  
GATTTAAAAGTTATTGATAAGACAGCTTTGCGGCGGCATATTAGCTATTTGCCGCAACA  
GGCCTATGTTTTTAGTGGCTCTATTATGGATAATCTCGTTTTAGGAGCTAAAGAAGGAA  
CGAGTCAGGAAGACATTATTCGTGCTTGTGAAATTGCTGAAATCCGCTCGGACATTGAA  
CAAATGCCTCAGGGCTATCAGACAGAGTTATCAGATGGTGCCGGTATTTCTGGCGGTCA  
AAAACAGCGGATTGCTTTAGCTAGGGCCTTATTAACACAGGCACCGGTTTTGATTCTGG  
ATGAAGCCACCAGCAGTCTTGATATTTTGACAGAAAAGAAAATTATCAGCAATCTCTTA  
CAGATGACGGAGAAAACAATAATTTTTGTTGCCACCGCTTAAGCATTTACAGCGGTAC  
TGACGAAGTCATTGTGTCATGGATCAGGGAAAAATTGTTGAACAAGGCACTCATAAGGAAC  
TTTTAGCTAAGCAAGGTTTCTATTATAACCTGTTTAAAT

➤ *S. mutans ComA* protein

MKQVIYVVLIVIAVNILLEIKRVTKRGGTVSSSNPLPDGQSKLFWRRHYKLVPQIDTR  
 DCGPAVLASVAKHYGSNYSIAYLRELSKTNKQGTALGIVEAAKKLGFETRSIKADMTL  
 FDYNDLTYPFIVHVIKGRKLQHYVVYGSQNNQLIIGDPDPSVKVTRMSKERFQSEWTG  
 LAIFLAPQPNYKPHKGEKNGLSNFFPLIFKQKALMTYIIIASLIVTLIDIVGSYYLQGI  
 LDEYIPDQLISTLGMITIGLIITYIIQQVMAFAKEYLLAVLSLRLVIDVILSYIKHIFT  
 LPMSFFATRRTGEITSRFTDANQIIDAVASTIFSIFLDMTMVILVGGVLLAQNNNLFFL  
 TLLSIPIYAIIFAFLEKPFKMNHEVMESNAVVSIIEDINGMETIKSLTSESARYQN  
 IDSEFVDYLEKNFKLHKYSAIQTALKSGAKLILNVVILWYGSRLVMDNKISVGQLITFN  
 ALLSYFSNPIENIINLQSKLQSARVANTRLNEVYLVESEFEKDGDLSENSFLDGDIFSE  
 NLSYKYGFGRDTLSDINLSIKKGSKVS LVGASGSGKTTAKLIVNFYEPNKGIVRINGN  
 DLKVIDKTALRRHISYLPQQAYVFGSIMDNLVLGAKEGTSQEDIIRACEIAEIRSDIE  
 QMPQGYQTELSDGAGISGGQKQRIALARALLTQAPVLILDEATSSLDILTEKKIISNLL  
 QMTEKTIIFVAHRLSISQRTDEVIVMDQKGKIVEQGTHKELLAKQGFYYNLFN

➤ *S. mutans* comB gene

ATGGATCCTAAATTTTTACAAAGTGCAGAATTTTATAGGAGACGCTATCATAATTTTGC  
 GACACTATTAATTGTTCTTTGGTCTGCTTGATTATCTTCTTGGTCATATTCCTTTGTT  
 TTGCTAAAAAAGAAATTACAGTGATTTCTACTGGTGAAGTTGCACCAACAAAGGTTGTA  
 GATGTTATCCAATCTTACAGTGACAGTTCAATCATTAAAAATAATTTAGATAATAATGC  
 AGCTGTTGAGAAGGGAGACGTTTTAATTGAATATTCAGAAAATGCCAGTCCAAACCGTC  
 AGACTGAACAAAAGAATATTATAAAAGAAAGACAAAACGAGAAGAGAAGGAAAAGAAA  
 AAACACCAAAAAGAGCAAGAAAAGAAGAAGTCTAAGAGCAAGAAAGCTTCCAAAGATAA  
 GAAAAAGAAATCGAAAGACAAGGAAAGCAGCTCTGACGATGAAAATGAGACAAAAAAGG  
 TTTTCGATTTTTGCTTCAGAAGATGGTATTATTCATACCAATCCCAAATATGATGGTGCC  
 AATATTATTCGAAGCAAACCGAGATTGCTCAAATCTATCCTGATATTCAAAAAACAAG  
 AAAAGTGTTAATCACCTATTATGCTTCTTCTGATGATGTTGTTTCTATGAAAAAGGGGC  
 AAACCGCTCGTCTTTCTTGGAAAAAAGGGAAATGACAAGGTTGTTATTGAAGGAAAA  
 ATTAACAATGTCGCTTCATCAGCAACTACTACTAAAAAAGGAAATCTCTTTAAGGTTAC  
 TGCCAAAGTAAAGGTTTCTAAGAAAAATAGCAAATCATCAAGTATGGTATGACAGGCA  
 AGACAGTCACTGTCATTGATAAAAAGACTTATTTTGATTATTTCAAAGATAAATTACTG  
 CATAAAATGGATAAT

➤ *S. mutans* ComB protein

MDPKFLQSAEFYRRRYHNFATLLIVPLVCLIIFLVIFLCFAKKEITVISTGEVAPTKVV  
 DVIQSYSDSSIIKNNLDNNAAVEKGDVLEIYSENASPNRQTEQKNI IKERQKREEKEKK  
 KHQSKKKKKSKSKKASKDKKKKSKDKESSDDENETKKVSI FASEDGI IHTNPKYDGA  
 NIIPKQTEIAQIYPDIQKTRKVLITYYASSDDVSMKKGQTARLSLEKKGNDKVIEGK  
 INNVAASATTTKGNLFKVTAKVKVSKKNSKLIKYGMTGKTVTVIDKKTYFDYFKDKLL  
 HKMDN

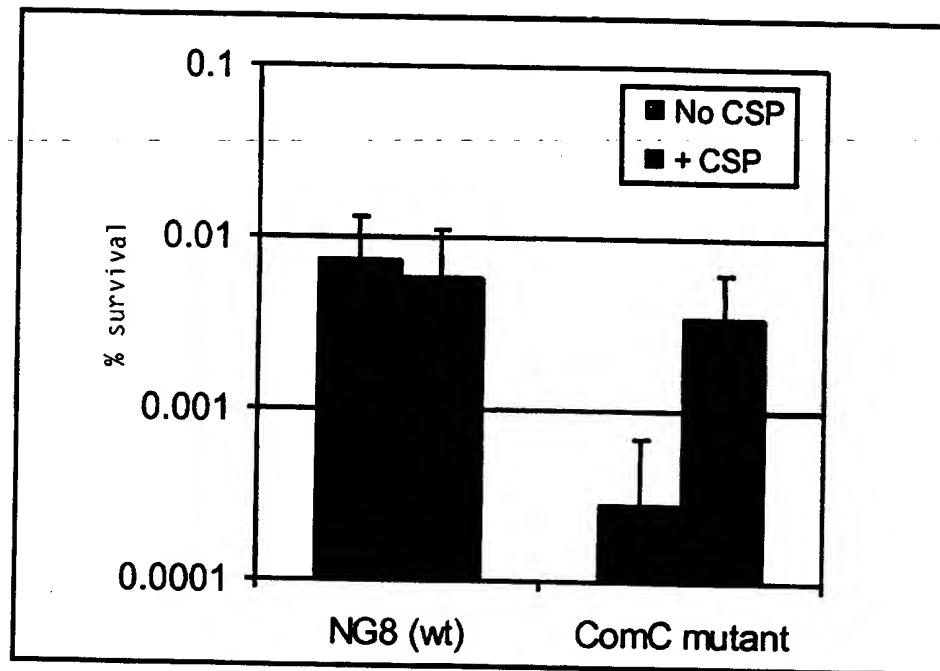


Figure 12